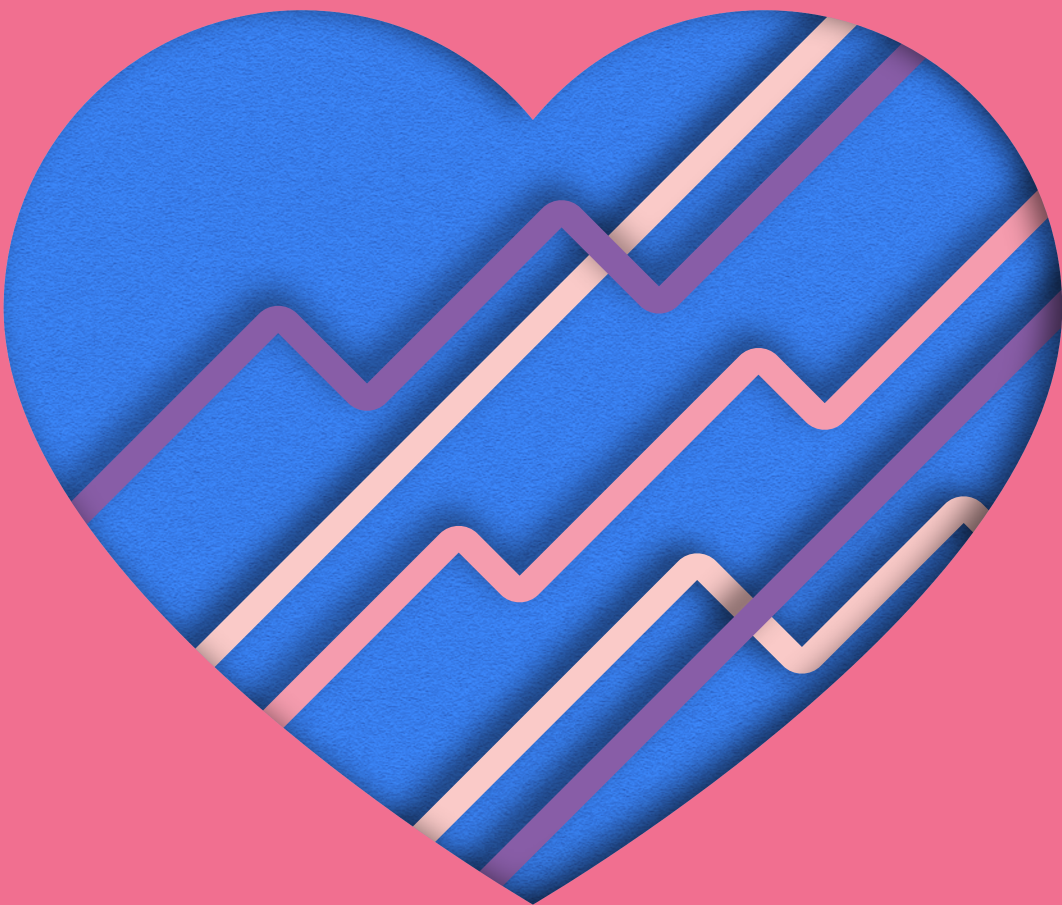
WORKING ON WELL-BEING

Using Empowerment and Behavioral Insights as Leadership Approaches to Improve the Well-being of Healthcare Employees



Henrico van Roekel

Working on Well-being

Using Empowerment and Behavioral Insights as Leadership Approaches to Improve the Well-being of Healthcare Employees

Henrico van Roekel

Working on Well-being: Using Empowerment and Behavioral Insights as Leadership Approaches to Improve the Well-being of Healthcare Employees.

ISBN: 978-94-6483-846-6
DOI: https://doi.org/10.33540/2221

Cover design: Bart de Koning ontwerp & Andrea Kroese | bartdekoning.nl
Internal design and print: Ridderprint | www.ridderprint.nl
This thesis was printed on 100% recycled paper.

Working on Well-being

Using Empowerment and Behavioral Insights as Leadership Approaches to Improve the Well-being of Healthcare Employees

Werken aan Welzijn

Empowerment en Gedragsinzichten Gebruiken als Leiderschapsbenaderingen om het Welzijn van Zorgmedewerkers te Verbeteren

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 31 mei 2024 des middags te 12.15 uur

door

Hendrik van Roekel
geboren op 1 juli 1995
te Ede

Promotoren:
Prof. dr. L.G. Tummers
Prof. dr. A.B. Bakker

Copromotor:
dr. C. Schott

Beoordelingscommissie:
Prof. dr. L. Bøgh Andersen
Dr. K. Breevaart
Prof. dr. S.M. Groeneveld
Prof. dr. E. Knies
Prof. dr. W.B. Schaufeli

Dit proefschrift werd mogelijk gemaakt met financiële steun van Stichting IZZ.

Contents

Chapter 1	*Working on Well-being*: Introduction	9
Chapter 2	Employee well-being during a crisis	25
Chapter 3	Text mining and employee well-being	35
Chapter 4	Empowering leadership during a crisis	73
Chapter 5	Shared leadership and employee willingness	99
Chapter 6	Motivation activation to increase ethical reporting	125
Chapter 7	Autonomy-preserving nudges to reduce email use	151
Chapter 8	Concluding *Working on Well-being*	177

Summary 198
Samenvatting in het Nederlands 206
References 214
List of preregistrations 242
Appendices 244
Co-author statements 292
Acknowledgements 298
About the author 300

Chapter 1

Working on Well-being: Introduction

1.1. Employee well-being in healthcare

'Health systems can only function with health workers.'
(World Health Organization, 2016, p.10)

'The line manager sets the mood and tone of the work environment and can therefore make or break a culture of well-being. As such, as a manager you are one of the most influential aspects in the workplace on a person's well-being, their discretionary effort and whether they stay in their role.'
(United Nations, 2022a, p.1)

Around the world and day-to-day, healthcare employees make invaluable contributions to the health and well-being of their patients. A recent World Health Organization report emphasized that healthcare employees are crucial in achieving the Sustainable Development Goals set by the United Nations (World Health Organization, 2016). Specifically, one central goal that will require healthcare employees' efforts is to 'ensure healthy lives and promote well-being for all at all ages' (United Nations, 2022b, overview section).

Healthcare employees are found in all sorts of workplaces: hospitals, nursing homes, home care, mental healthcare, and disability care. They hold various jobs, such as doctor, nurse, caretaker, paramedic, psychologist, social worker, HR employee, or manager. Take the Netherlands, with more than 1.4 million healthcare employees. They make up 17% of the working population (CBS, 2022a). These employees appear to do a respectable job, as the Dutch healthcare system scores high in rankings (e.g., Schneider et al., 2021). Compared to all OECD countries, Dutch citizens, together with the Norwegian, Belgian, and Swiss, are most likely to be satisfied with the availability of quality health services (92% are satisfied; OECD, 2021, p.126). Less than 0.5% of the Dutch population reported unmet medical care needs (OECD, 2021, p.128). Not coincidentally, population health also scores high: #15 out of 169 countries, according to the Bloomberg Global Health Index (Miller and Lu, 2019).

Healthcare employees work hard to improve the well-being of others. However, their well-being has increasingly been put under pressure. But how and why? We use Job Demands-Resources (JD-R) theory to explain this (Bakker et al., 2023). The core argument of JD-R theory is that employees are confronted with two types of job characteristics: job demands, those aspects of a job that cost energy, and job resources, those aspects of a job that enrich and motivate. Job demands instigate a health impairment process that leads to exhaustion and burnout, whereas job resources start a motivational process that leads to work engagement (Bakker and Demerouti, 2017; Schaufeli et al., 2009). Employees who experience burnout are chronically exhausted and have cynical attitudes toward their work. In contrast, work engagement is a work-related, positive state of mind, in which employees experience vigor, dedication, and absorption (Bakker et al., 2014; Schaufeli and Bakker, 2010).

Studies have identified a range of job demands that lead to burnout among healthcare employees. For example, burnout among physicians relates to work (e.g., long working hours), personal (e.g., over-commitment), and organizational characteristics (e.g., toxic leadership) (Patel et al., 2018). A Dutch study showed that, in comparison to other sectors, healthcare employees are more likely to experience high job demands, like high emotional burden, as well as low job resources, like a lack of autonomy (TNO, 2020). As a result, healthcare worker burnout is prevalent across many countries, healthcare systems, and job types, although estimates about the percentage of burnout vary (for example, between 6% and 47% among intensive care unit professionals; Chuang et al., 2016). A recent survey in the Netherlands found that one out of six healthcare employees report they are 'often or always' emotionally exhausted, whereas one out of five employees are physically exhausted (Van der Fels, 2022).

Burnout among healthcare employees can result in a range of adverse health problems, including anxiety, depression, and higher suicide rates (Patel et al., 2018; Yang and Hayes, 2020). Healthcare employees are also more likely to get occupational diseases compared to other sectors, such as cardiovascular diseases and musculoskeletal disorders (TNO, 2020; Toppinen-Tanner et al., 2009). Besides, burnout is negatively related to healthcare employee performance. It is associated with, for instance, increased rates of medical mistakes (Wen et al., 2016). Likewise, absenteeism is higher for healthcare employees in comparison to other sectors, especially among nurses and caretakers (TNO, 2020). The above statistics decrease the attractiveness of healthcare as an employer (e.g., Tummers et al., 2013): four out of ten Dutch healthcare employees consider leaving the healthcare sector (Van der Fels, 2022). A study that generated labor market prognoses for the healthcare sector indicated that in the coming years, the shortage of healthcare employees will continue to grow, especially in nursing homes and home care (ABF Research, 2022).

All these statistics seem to suggest that working in healthcare is not very healthy. Societal developments paint a complex picture of why healthcare employees' job demands and resources have been severely affected. First, the worldwide economic crisis in the late 2000s heralded an era of budget cuts and reforms within healthcare systems worldwide (International Labour Office, 2013). Second, the increasing focus on efficiency, productivity, and measurement can already be traced back to the introduction of New Public Management logic in healthcare sectors across Europe in the 1980s (Hood, 1991; Noordegraaf and Abma, 2003). Third, society is aging. Not only is there an aging patient population, but the healthcare workforce is also getting older which renders labor an increasingly scarce resource (Van Dalen et al., 2010; Vonk et al., 2020). Fourth, technical innovations in medical treatments have increased healthcare usage and expenditure and will continue to do so for the coming decades (Vonk et al., 2020). Fifth, the COVID-19 pandemic has impacted the well-being of healthcare employees (e.g., Conti et al., 2021). Healthcare employees experienced the burden of an increasingly ill population (Spoorthy, Pratapa and Mahant, 2020).

To conclude, improving healthcare employee well-being is a big challenge for our society. Healthy healthcare employees are a crucial precondition for the health and well-being of patients and clients. How can we take care of healthcare employees? In this thesis, we aim to deepen our understanding of healthcare employee well-being by studying the factors that affect well-being and developing strategies that could help improve well-being (e.g., Kniffin et al., 2021).

1.2. Employee well-being and leadership

The extant literature has identified a variety of resources that can positively impact employee well-being. For example, studies show that situational resources—such as autonomy and social support—and personal resources such as self-efficacy and optimism—are positively associated with work engagement (Bakker et al., 2023; Lesener et al., 2020; Mäkikangas et al., 2013). One higher-order factor and critical resource that we focus on in this research concerns leadership.

Leadership is traditionally defined as the process of social influence towards goals, enacted by people with formal leadership roles in the organization towards their followers (Antonakis and Day, 2017; Kelloway and Dimoff, 2017). To understand the influence that leaders have on employees, leadership theory and research have dramatically expanded in the past decades (Dinh et al., 2014). Besides the emergence of a distinct leadership field, the study of leadership has penetrated many disciplines—like public administration (‘t Hart and Tummers, 2019), and specific occupational sectors—like healthcare (e.g., Sfantou et al., 2017). We now know that leadership behaviors can affect employees’ job demands and resources and, consequently, employee well-being. Based on a literature review, Tummers and Bakker (2021) identified three ways in which leadership can positively impact employees' job demands and resources. First, leaders can increase job resources (e.g., increase social support or development opportunities) or decrease job demands (e.g., reduce workload). Second, leaders can influence the strength of the links between job demands and resources and their outcomes. For example, by increasing employees' autonomy leaders can reduce the impact of job demands on job strain, even if leaders cannot decrease the job demands themselves. Third, leaders can influence how employees deal with resources by encouraging job crafting, which refers to the process in which employees proactively change their jobs to increase the person-job fit (Tims et al., 2015).

Besides this instrumental approach to leadership, scholars have studied how certain leadership styles affect well-being. For example, transformational leadership describes how leaders can be visionary and creative to inspire employees, which is generally associated with higher well-being (Arnold, 2017; Breevaart and Bakker, 2018). In contrast, leaders can also negatively affect well-being. Destructive leadership refers to a leadership style that includes well-being-undermining behaviors such as bullying. This leadership style is negatively associated with employee work engagement and positively associated with employee burnout (Breevaart et al., 2014; Pletzer et al., 2023).

Compelled by the challenge of healthcare employee well-being, the search for approaches towards leadership to support employee well-being is increasingly important. In this dissertation, we aim to contribute to finding such approaches to leadership. Rather than reevaluating traditional approaches to leadership, we study how two contemporary approaches can contribute to *Working on Well-being.* They are contemporary because they respond to findings on human cognition and abilities in the last decades. Moreover, they focus on concrete behaviors that leaders can display, which increases the potential for practical application. However, the approaches have different foundations and are rooted in different disciplines. The approaches we consider are empowerment and behavioral insights.

1.2.1. Empowerment

Within the management and organizational behavior literature, scholars have observed how a more skilled and knowledgeable workforce has increasingly led leaders to adopt a more emancipatory view of leadership (Bartunek and Spreitzer, 2006; Parker, Wall and Corderly, 2001). In this view, leaders can empower employees. In a work context, empowerment refers to the transfer of influence and power from formal leaders to employees (Amundsen and Martinsen, 2014). Below, we discuss two leadership styles that are key to this approach: empowering leadership and shared leadership.

First, empowering leadership refers to a leadership style in which leaders, according to Ahearne et al. (2005), enhance the meaningfulness of work, foster participation in decision-making, express confidence in high performance, and provide autonomy from bureaucratic constraints. Antecedents of empowering leadership include low power distance between leaders and employees and high leaders' perceptions of team capability (Tang et al., 2020). In turn, empowering leadership provides employees with job resources like autonomy and psychological capital (Park et al., 2017). Consequently, empowering leadership is associated with increases in employee well-being, like work engagement and job satisfaction (Kim et al., 2018), and performance (Lee et al., 2018). However, some scholars have pointed out that empowerment may have a dark side by, for example, increasing the burden placed on employees (e.g., Cheong et al., 2016).

Second, shared leadership describes how leadership within a team or organization is dispersed among its members. Scholars generally agree that three characteristics define shared leadership: it refers to lateral influence between peers, emerges within a team, and implies the dispersion of roles and influence among its members (Zhu et al., 2018). Antecedents of shared leadership that have been found include an environment in which employees have a voice, show support, and express mutual trust (Carson et al., 2007; Drescher et al., 2014). Likewise, the formal leader should empower and motivate employees and be humble for employees to be able to share leadership (Chiu et al., 2016; Fausing et al., 2015; Hoch, 2013). As a result, shared leadership provides employees with job resources like collective efficacy and team cohesion (Zhu et al., 2018). It is also positively associated with team satisfaction, creativity, and performance (D'Innocenzo et al., 2016). Besides, shared leadership is associated with decreases in leader stress, as their central position of power diminishes (Foulk et al., 2018).

Consider this example of what empowerment could look like in healthcare organizations. Team leaders, among other tasks, perform leadership behaviors aimed at relations, such as supporting individual team members (Yukl, 2012). Team leaders hereby solely have a position of influence. However, employees could also assume some of the leadership tasks aimed at relations. For example, an employee could become the team counsellor.

1.2.2. Behavioral insights

Next to empowerment, we consider behavioral insights. These insights from the behavioral sciences are based on empirical results that describe how humans truly make choices. Within behavioral economics, Simon (1955) suggested that people are less optimal decision-makers than economists often expected. It has been found that human decision-making is subject to many heuristics, which are mental shortcuts that can lead to biased decisions (Tversky and Kahneman, 1974). This is caused by what Hallsworth and Kirkman (2020, p.2) called the core behavioral insight: 'Much of our behavior is nonconscious, habitual, and driven by cues in our environment or how choices are presented.' This observation questions the existence of the homo economicus—the notion that humans are entirely rational decision-makers. Instead, people are boundedly rational: they make decisions that they believe are good enough (i.e., satisficing) rather than perfect decisions (i.e., optimizing) (Simon, 1955; Tversky and Kahneman, 1981).

Behavioral insights propose alternative ways of influencing behavior to traditional managerial instruments. By taking biases and flaws that human decision-making has into account, behavioral insights can inform the way leaders influence employees. In the book *Nudge,* first published in 2008, Thaler and Sunstein (2021) described how nudges can help employees make better decisions. A nudge is 'any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives' (ibid., p.8). It aims to influence behavior by making use of insights on bounded rationality (Hansen, 2016). Nudges can work through a variety of mechanisms, including changing defaults, giving reminders or providing social reference points (Münscher et al., 2016). An example of the latter is described by Hallsworth et al. (2016), who decreased antibiotics prescriptions by showing overprescribing general practitioners that they prescribed more than their peers.

In recent years, the literature on behavioral insights has rapidly developed. Within public administration, a sub-discipline devoted to studying behavioral insights in a public context emerged, called behavioral public administration (Grimmelikhuijsen et al., 2017). Here, behavioral insights are used to study, for example, public service motivation: civil servants' desire to contribute to society (Meyer-Sahling et al., 2019). However, multiple authors have argued that the literature still needs to mature (Bhanot and Linos, 2020; Hassan and Wright, 2020). For example, there is considerable academic discussion regarding the definitions of nudges (Hansen, 2016), the extent to which they are desirable (Wachner et al., 2021; Wilkinson, 2013), and the extent to which nudges can cause behavior change (Maier et al., 2022; Mertens et al., 2022). Additionally, scholars have also developed alternative behavioral interventions like boosts and self-nudges (Hertwig and Grüne-Yanoff, 2017; Reijula and Hertwig, 2022).

There is little research regarding the use of behavioral insights in the workplace and the relationship with employee well-being in healthcare. Nevertheless, some studies have suggested that leaders could employ behavioral insights, like nudges, to improve employee well-being. For example, Weintraub et al. (2021) used goal-setting nudges to improve flow, thereby decreasing stress and improving work engagement. Georganta and Montgomery (2016) suggested nudges could be used to increase workplace fun, which they conceptualize as an essential job resource. Nudges may, for example, increase team cohesion and collaboration. This shows that interventions could improve employee well-being by targeting concrete outcomes such as changing employee behaviors (e.g., making it easier for employees to work more safely) or reducing the stress involved in decisions employees must make (e.g., alleviating cognitive pressure through reminders) (see e.g., Nagtegaal et al., 2019).

Consider this example about how behavioral insights may increase healthcare employee well-being. Studies show that email at work can cause stress, especially if employees feel it is necessary to read and respond to emails in real time (Brown et al., 2014; Giurge and Bohns, 2021). To help healthcare employees deal with this, leaders could communicate that they are going to email less to reduce stress. This does not only directly reduce the emails sent, but most importantly the message functions as an opinion leader nudge (Münscher et al., 2016). Being presented with a descriptive social norm—a statement about what other people would do in a situation—from a respected messenger like the team leader, people are more likely to conform (Valente and Pumpuang, 2007).

Table 1 summarizes both approaches to leadership introduced above by discussing their aim, the discipline in which they evolved, the concepts that they use, and an example.

Table 1 *Two approaches to leadership: empowerment and behavioral insights*

	Empowerment	**Behavioral insights**
Aim	To transfer influence to employees	To ease employees' decision-making
Discipline	Management; Organizational behavior	Behavioral sciences; Behavioral economics; Behavioral public administration
Concepts	Empowering leadership; shared leadership	Behavioral interventions; nudges; choice architecture
Example	An employee becomes the team counsellor (chapter 5)	The leader communicates that they are going to email less, which nudges employees to follow (chapter 7)

1.3. Research questions and contributions to theory

In this dissertation, we aim to consider how leadership can contribute to employee well-being in healthcare. This is a complex challenge that involves multiple disciplines, theories and concepts. To do this challenge justice, we do not aim to explore one single phenomenon in-depth. Instead, we aim to advance our understanding of leadership and employee well-being by developing studies that address a variety of topics and display a breadth of possible approaches and tools. The studies are distinct research projects but share this common goal.

Three main research questions guide our process of studying *Working on Well-being.* Each of the questions is answered through two studies. For the first research question, we turn our attention to employee well-being itself. While we observed that healthcare employee well-being is increasingly being put under pressure, our understanding of it is limited in several ways. Before we evaluate the potential of empowerment and behavioral insights, we first ask:

RQ1: *How can we deepen our understanding of employee well-being in healthcare?*

To answer this question, we found two literature gaps that we can address to improve our understanding of employee well-being: differentiation between groups of healthcare employees and the innovation of well-being measurement.

First, we found that recent studies in a Dutch context that measure employee well-being differentiate little between groups of healthcare employees (Shreffler et al., 2020; TNO et al., 2020). While general statistics about healthcare employee well-being may give reason to worry, we know little about whether specific groups of healthcare employees experience lower well-being than others (Shreffler et al., 2020). We therefore aim to make an empirical contribution to the literature on healthcare employee well-being by studying how the experiences of Dutch healthcare employees differed in a specific context: during the recent COVID-19 crisis. **Chapter 2** presents the results of a cross-sectional survey among 7,208 healthcare employees in which we study the extent to which healthcare employees dealing with COVID-19 patients reported lower well-being on several indicators. Besides, we explore what personal and work characteristics are associated with lower employee well-being.

The second limitation that is prevalent in the literature on employee well-being is that studies are overly dependent on traditional methods of measurement. Most studies use validated scales to measure employee well-being (Bakker et al., 2014). Such scales are limited in, for example, presenting the multidimensionality of a phenomenon and allowing for new theoretical discoveries (Balducci and Marinova, 2018). By evaluating a novel approach to measure well-being, we could critically assess the extant literature on employee well-being and find out to what extent such an alternative method confirms, extends, and questions theory and findings (Jurafsky and Martin, 2017; Kobayashi et al., 2021). This would increase both the rigor and societal relevance of well-being research. **Chapter 3** introduces text mining as a novel method to measure well-being. We focus on a specific employee well-being construct: work engagement. The goal of our study is to explore whether we can use text mining to classify healthcare employees into high or low work engagement and analyze the specific text features that contribute to classification. We use self-written narratives of healthcare employees in two surveys (n = 5,591 and n = 4,470).

After having contributed to the literature on employee well-being in these two distinct ways, we turn to the first of the leadership approaches. Our second research question is:

RQ2: *How can leaders use empowerment to contribute to employee well-being in healthcare?*

To answer this question, we analyzed the literature on the two leadership styles often studied in this context: empowering leadership (Kim et al., 2018; Lee et al., 2018) and shared leadership (Zhu et al., 2018). While we know a lot about the antecedents and outcomes of both styles, there are pressing literature gaps. For empowering leadership, the role of context is understudied. For shared leadership, we know little about employees' personal preferences.

First, regarding empowering leadership, the literature generally agrees that it increases employee well-being (Kim et al., 2018). However, some studies showed that empowering leadership can have dark sides when it, for example, increases the burden for employees (Cheong et al., 2016). These conflicting findings may be caused partially by neglecting how effects differ depending on the context in which they appear: ignoring context is a returning criticism of leadership studies (Kim et al, 2018; Sims et al., 2009). We aim to address this gap by studying the effect of context in the relationship between empowering leadership and employee well-being. Specifically, we investigate whether the effect of empowering leadership on employee well-being differs during a public health crisis ('t Hart and Tummers, 2019; Zhou et al., 2020). Such a study may help explain the contrasting literature on the two faces of empowering leadership (e.g., Cheong et al., 2016), and it provides an empirical test of the effects of empowerment in crisis leadership ('t Hart and Tummers, 2019). **Chapter 4** presents a natural experiment in which we combine a longitudinal survey (n = 468) with administrative data. We exploit the geographical variance in COVID-19 hospitalization rates to study whether the effect of empowering leadership on healthcare employee well-being differs depending on the intensity of a public health crisis.

Secondly, in the literature on shared leadership, scholars have studied the necessary preconditions for shared leadership to arise and their consequences for leaders and employees (e.g., Carson et al., 2007; Drescher et al, 2014). Whereas most of the literature focused on higher-level antecedents on the team or formal leader level—such as a leader's levels of empowering leadership, we contribute by approaching shared leadership from the bottom-up perspective of employees (Zhu et al., 2018). What types of shared leadership do employees want to execute? By asking this question, we aim to open the black box of personal considerations in shared leadership emergence and improve our understanding of shared leadership's opportunities and pitfalls for organizations and employees (Jønsson et al., 2016; Yukl, 2002). **Chapter 5** presents a conjoint experiment in which 6,742 healthcare employees assess which shared leadership behaviors they would execute. We also analyze how preferences vary across personal characteristics—such as gender and age.

Having evaluated how empowerment could be part of *Working on Well-being,* our third and final research question addresses the second leadership approach:

RQ3: *How can leaders use behavioral insights to contribute to employee well-being in healthcare?*

There are only a few studies that explore how leaders could employ behavioral insights to improve employee well-being in general and in healthcare (e.g., Weintraub et al., 2021). How can we improve our understanding of the potential of behavioral insights for well-being? First, scholars have urged us to study the mechanisms of behavior change. Second, they have suggested paying more attention to the pros and cons of behavioral interventions in field settings. We aim to address those issues and we do so in the context of behaviors related to healthcare employee well-being.

First, within behavioral public administration, scholars have urged to go beyond quick wins by rigorously studying mechanisms of behavioral change (Bhanot and Linos, 2020). We draw on a recent experimental study that showed that activating employees' public service motivation can increase the ethical reporting of wrongdoers (Meyer-Sahling et al., 2019). Herein, public service motivation describes the desire to contribute to society (Perry and Hondeghem, 2008). We aim to study this phenomenon more. First, we ask whether other concepts can do the same. For example, prosocial motivation describes the desire to help others (Grant, 2008a). Next, we aim to test whether effects differ for different levels of motivation or different wrongdoers. Studying the mechanisms of behavioral change, such as how and when motivations can be activated and effective, presents insight into the potential of motivation activation as a behavioral intervention. Such an intervention may improve employee well-being, as wrongdoings in organizations affect employee well-being negatively (Belle and Cantarelli, 2017; Ripoll, 2019; Searle and Rice, 2020). **Chapter 6** presents a study in which we investigate whether we can activate the motivations of healthcare employees to increase ethical reporting, utilizing a question-order survey experiment among 11,728 employees.

Second, scholars have suggested that the pros and cons of behavioral interventions such as nudges should be assessed more critically (Bhanot and Linos, 2020) and more (quasi) field experiments should be used to test the promises of behavioral insights in practice (Hassan and Wright, 2020). Nudges are popular behavioral interventions that leaders could use to influence employees' behaviors, but nudges have also been critiqued for harming the autonomy of employees (Hausman and Welch, 2010; Wilkinson, 2013), and for often being ineffective (Maier et al., 2022; Szaszi et al., 2022). In this study, we aim to address these criticisms by studying how leaders could use innovative nudges that are autonomy-preserving and effective (e.g., Reijula and Hertwig, 2022) and assessing their effectiveness in both survey and field settings. **Chapter 7** reports how we attempt to develop nudges that are autonomy-preserving and effective in reducing email use among healthcare employees (Smith and Lewis, 2011). As email use is often a source of stress, nudges that reduce email use may contribute to employee well-being (Brown et al., 2014; Reinke and Chamorro-Premuzic, 2014). We use a qualitative pre-study to develop nudges, a pilot survey (n = 435) and a survey experiment (n = 4,112) to test perceived autonomy and subjective effectiveness, and a quasi-field experiment (n = ±1,189) to test objective effectiveness.

1

1.4. Dissertation overview

The following six chapters each present one of the studies conducted. Figure 1 presents an overview of all the chapters. Table 2 presents the research question, methods and publication status of the article-based chapters. In our last chapter, chapter 8, we conclude and discuss implications for theory, methods and practice.

Figure 1 *Overview*

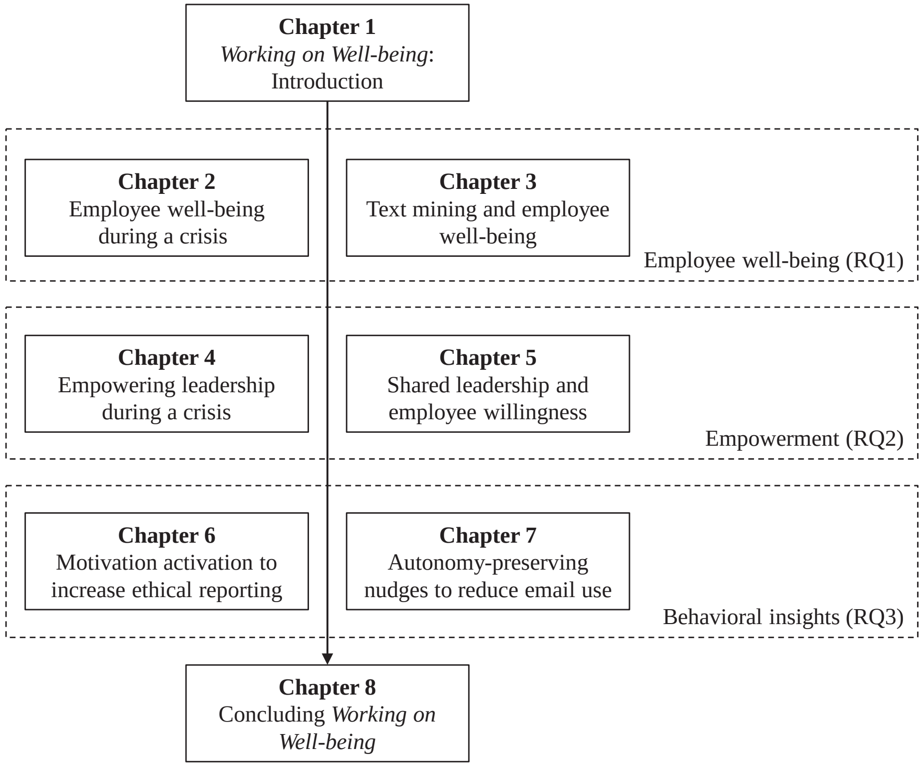


Table 2 *Summary*

Chapter	Research question	Methods	Publication status
RQ1: Employee well-being			
2	Have healthcare workers dealing with COVID-19 patients experienced more threats to well-being than other healthcare workers, and are there differences within the group of healthcare workers who work with COVID-19 patients?	Cross-sectional survey (n = 7,208)	Published in *Frontiers in Psychology*
3	Can we explain work engagement by analyzing self-narratives through text mining?	Text mining of self-narratives across two surveys (n = 5,591 and n = 4,470)	Published in *Applied Psychology: An International Review*
RQ2: Empowerment			
4	Does the effect of empowering leadership on employee well-being differ during crisis?	Natural experiment with longitudinal survey and administrative data (n = 468)	*Under review*
5	Which shared leadership behaviors do employees want to exercise?	Conjoint survey experiment (n = 6,742)	Published in *Leadership*
RQ3: Behavioral insights			
6	Does the activation of public service motivation and prosocial motivation influence the intentions to report wrongdoings from colleagues differently compared to the intentions to report wrongdoings from patients?	Question order survey experiment (n = 11,728)	Published in *Public Management Review*
7	To what extent can nudges preserve autonomy and be effective in decreasing email use?	Qualitative pre-study, pilot survey (n = 435), survey experiment (n = 4,112), quasi-field experiment (n = ±1,189)	Published in *Behavioural Public Policy*

Note. Due to specific journal requirements, chapters may vary slightly in structure or style.

Chapter 2

Employee well-being during a crisis

This chapter is based on the following published article:
Van Roekel, H., van der Fels, I. M., Bakker, A. B., & Tummers, L. G. (2021). Healthcare workers who work with COVID-19 patients are more physically exhausted and have more sleep problems. *Frontiers in Psychology, 11*. DOI: 10.3389/fpsyg.2020.625626.

Abstract

In this survey study of 7,208 Dutch healthcare workers, we investigate whether healthcare workers dealing with COVID-19 patients experience lower general health, more physical and mental exhaustion and more sleep problems than other healthcare workers. Additionally, we study whether there are differences in well-being within the group of healthcare workers working with COVID-19 patients, based on personal and work characteristics. We find healthcare workers who are in direct contact with COVID-19 patients report more sleep problems and are more physically exhausted than those who are not in direct contact with COVID-19 patients. Mental exhaustion and general health do not significantly differ between healthcare workers who are in direct contact with COVID-19 patients and those who are not. Among healthcare workers in direct contact with COVID-19 patients, lower well-being on one or more indicators is reported by those who are female, living alone, without leadership role, or without sufficient protective equipment. Regarding age, physical exhaustion is more prevalent under healthcare workers older than 55 years, whereas mental exhaustion is more prevalent under healthcare workers younger than 36 years. These results stress the need of mental and physical support of healthcare workers during a pandemic, catered to the needs of healthcare workers themselves.

2.1. Introduction

The COVID-19 pandemic has presented great threats to the well-being of healthcare workers. Many of them risked infection with the virus while working longer hours in understaffed organizations (Adams and Walls, 2020; Mhango et al., 2020; Pearman et al., 2020; Wang, Zhou et al., 2020). Since the outbreak, scholars have presented first results on what effects the crisis has had on healthcare workers. Studies show effects on attitudes and practices, like a high fear of self-infection (Zhou et al., 2020), an increase in mental health problems like job stress and anxiety (Cao et al., 2020; Spoorthy et al., 2020; Tan et al., 2020; Wei et al., 2020), and the development of physical problems like increased headaches due to wearing protective equipment (Ong et al., 2020). Similarly, a scoping review of 37 studies on how COVID-19 has impacted healthcare worker wellness showed COVID-19 was associated with, among else, more stress, anxiety and poorer quality of sleep (Shreffler et al., 2020).

However, we know little about whether the effects of the COVID-19 outbreak on healthcare workers' well-being differ across groups of healthcare workers. We therefore firstly study whether healthcare workers dealing with COVID-19 patients experience more threats to well-being than other healthcare workers. For instance, is it truly the case that healthcare workers working with COVID-19 patients report more exhaustion? Second, we study whether there are differences within the group of healthcare workers who work with COVID-19 patients. Besides, studies on healthcare worker well-being are mainly conducted in Asian context (Cao et al., 2020; Shreffler et al., 2020; Spoorthy et al., 2020; Tan et al., 2020; Zhou et al., 2020). We present data on Dutch healthcare workers to address these gaps.

As healthy healthcare workers are crucial in the aftermath of the outbreak, and in prevention of further outbreaks, losing a substantial part of the workforce to psychological or physical threats is detrimental. Therefore, the results can fuel healthcare organization policies and human resource practices to sustain the mental and physical health of healthcare workers during and after COVID-19.

2.2. Methods

We collected data in a May-June 2020 cross-sectional survey on work and health of Dutch healthcare workers[1]. Healthcare workers were invited via email to voluntarily participate in the online survey and they were reminded after a few weeks. To protect their identities, respondents were not asked to give their names and contact information; other potentially identifiable data, such as gender, age, and job type were carefully protected. A total of 7,208 respondents completed our survey. Data used in this article are included as an appendix.

We use four employee well-being measures as dependent variables: a general health measure asking respondents to rate their general health (10-point scale ranging from 1 to 10 (Sullivan and Karlsson, 1998)), mental exhaustion (five items on a 5-point Likert scales ranging from 1 (never) to 5 (always) (daily), example item: I feel mentally exhausted because of my work (Schaufeli et al., 1996)), physical exhaustion (five items on a 5-point Likert scales ranging from 1 (never) to 5 (always) (daily), example item: I feel physically exhausted because of my work (Schaufeli et al., 1996)), and sleep problems (three items on a 5-point Likert scales ranging from 1 (no) to 5 (a lot), example item: I have a restless or disturbed sleep (Adriaenssens et al., 2012)).

For our independent variables, we compare the well-being outcomes between groups based on personal and work characteristics. First, we assess whether outcomes differ for healthcare workers who do and do not work in direct contact with COVID-19 patients. Next, within the group of healthcare workers who work with COVID-19 patients, we assess multiple variables to define risk groups of healthcare workers. To do so, we study three personal characteristics: gender, age, and whether the healthcare worker lives alone. For age, we divide our sample into three categories: younger than 36, between 36 and 55, and older than 55. This is a common division of younger, middle-aged and older employees used in academic research as well as governmental research on well-being. It also enables to assess non-linear relationships with well-being. Additionally, we study two important work characteristics: leadership role (whether the healthcare worker indicates to have a leadership role) and sufficient protective equipment (healthcare workers were asked: ‘do you have sufficient protective equipment at your disposal?’; they could answer with yes or no). In selecting these variables, we have not aimed to be exhaustive, but to constitute a broad picture of factors potentially related to well-being.

1 The studies involving human participants were reviewed and approved by the Faculty Ethical Review Committee of Utrecht University. The participants provided their written informed consent to participate in this study.

Our sample (n = 7,208) is representative for Dutch healthcare workers in terms of gender: our sample has 82% females, while for Dutch healthcare workers this is 84%. However, our sample is older (M = 51.5 versus M = 42.5) (CBS data[2]). Furthermore, our respondents represent all healthcare industries: hospitals (36.2%), nursing homes and homecare (23.6%), mental health care (16.5%), disability care (17%) and other healthcare industries (6.7%).

For analyses, we conduct t-tests or ANOVAs, when appropriate. For the ANOVAs we conduct post hoc analyses (Tukey's HSD) to define which groups significantly differ. The level of significance is set at 0.05 and Cohen's *d* effect sizes are calculated (Cohen, 1988). Secondly, as additional analysis, multivariate regression analyses are performed for each of the four well-being variables to gain more understanding on the relative strength with which the variables are related to well-being. The defined groups are included as independent variables. We report adjusted R-squared values for the models and Beta-values to indicate the relative strength of each variable.

2.3. Results

We start by contrasting healthcare workers who work in direct contact with COVID-19 patients versus those who do not (Table 1). Healthcare workers in direct contact with COVID-19 patients report significantly more sleep problems and physical exhaustion. No significant differences are found for mental exhaustion or general health.

Next, we zoom in within the group of healthcare workers in direct contact with COVID-19 patients (Table 2). First, female healthcare workers report more sleep problems and physical exhaustion than male healthcare workers, whilst there are no significant differences on mental exhaustion and general health.

2 https://azwstatline.cbs.nl

Table 1 *More sleep problems and physical exhaustion for healthcare workers in direct contact with COVID-19 patients*

Direct contact COVID-19 patients	N	Sleep problems			Physical exhaustion			Mental exhaustion			General health		
		***M* (*SD*)**	***t*(7,206)**	***d***	***M* (*SD*)**	***t*(7,206)**	***d***	***M* (*SD*)**	***t*(7,206)**	***d***	***M* (*SD*)**	***t*(7,206)**	***d***
Yes	2,621	2.42 (.97)	-6.45**	.159	2.29 (.82)	-9.21**	.223	1.98 (.76)	-0.64	-	7.61 (1.26)	-1.49	-
No	4,587	2.27 (.91)			2.11 (.79)			1.97 (.74)			7.56 (1.31)		

Note. Cohen's *d* effect sizes are small (Cohen, 1988). ** $p < 0.01$.

Table 2 *Differences within the group of healthcare workers who are in direct contact with COVID-19 patients*

		Sleep problems			Physical exhaustion			Mental exhaustion			General health		
	N	*M* (*SD*)	*t*(2,619)	Cohen's *d*	*M* (*SD*)	*t*(2,619)	Cohen's *d*	*M* (*SD*)	*t*(2,619)	Cohen's *d*	*M* (*SD*)	*t*(2,619)	Cohen's *d*
Gender													
Female	2,201	2.46 (.97)	4.73**	.226	2.31 (.81)	2.73*	.145	1.98 (.76)	.41	-	7.60 (1.24)	-.59	-
Male	420	2.21 (.91)			2.19 (.84)			1.96 (.77)			7.64 (1.40)		
Age													
< 36 years	277	2.36 (.96)	1.09[c]	-	2.30 (.80)	6.77**[c]		2.10 (.83)^	4.41*[c]	.191[b]	7.61 (1.21)	.49[c]	-
36 - 55 years	1,347	2.41 (.97)			2.23 (.79)			1.95 (.74)			7.63 (1.26)		
> 55 years	997	2.45 (.96)			2.36 (.86)^		.157[a]	1.98 (.78)			7.58 (1.29)		
Living alone													
No	2,151	2.41 (.97)	-1.24	-	2.26 (.81)	-3.12**	.158	1.96 (.75)	-3.18**	.155	7.64 (1.25)	2.20*	.112
Yes	470	2.47 (.95)			2.39 (.84)			2.08 (.80)			7.49 (1.32)		
Leadership role													
No	2,199	2.43 (.96)	1.20	-	2.30 (.82)	2.51*	.121	1.98 (.76)	.36	-	7.59 (1.26)	-1.75	-
Yes	422	2.37 (.99)			2.20 (.83)			1.97 (.77)			7.71 (1.31)		
Sufficient protective equipment													
No	507	2.67 (1.03)	-6.67**	.314	2.60 (.89)	-9.90**	.466	2.33 (.84)	-11.98**	.562	7.24 (1.40)	7.48**	.352
Yes	2,114	2.36 (0.94)			2.21 (.78)			1.89 (.72)			7.70 (1.21)		

Note. Cohen's *d* effect sizes are small to medium (Cohen, 1988). * $p < 0.05$; ** $p < 0.01$. ^ significantly different from healthcare workers between 36 and 55 years; [a] effect size between age categories 36-55 years and >55 years; [b] effect size between age categories <36 years and 36-55 years; [c] *F* -test ($df = 2{,}618$). Regarding healthcare workers' age, physical exhaustion is more prevalent among healthcare workers who are older than 55 compared to healthcare workers between 36 and 55 years old. In contrast, mental exhaustion is more prevalent among healthcare workers who are younger than 36, compared to healthcare workers between 36 and 55 years old. There are no significant differences between age categories on sleep problems and general health.

2

Additionally, we assess whether living alone or with family is correlated with well-being. We find that healthcare workers who live alone report higher physical and mental exhaustion and lower general health. No significant differences are found for sleep problems.

Next, we consider work characteristics. Healthcare workers without a leadership role are found to be more physically exhausted than healthcare workers who have a leadership role. No significant differences of having a leadership role are found for sleep problems, mental exhaustion and general health.

Finally, is having sufficient protective equipment in working with COVID-19 patients correlated with well-being? We find significant differences for all outcomes: healthcare workers who do not have sufficient protective equipment report more sleep problems, more physical and mental exhaustion, and lower general health.

In additional analysis we conduct multivariate regression analyses per well-being outcome. The analyses yield similar results as above. For sleep problems, gender (reference = female; $\beta = -0.09$, $t(2{,}614) = -4.49$, $p < 0.05$) and having sufficient protective equipment (ref. = sufficient equipment; $\beta = 0.12$, $t(2{,}614) = 6.35$, $p < 0.05$) are significant predictors (Adj. $R^2 = 0.024$). For physical exhaustion, gender ($\beta = -0.05$, $t(2{,}614) = -2.40$, $p < 0.05$), living alone (ref. = not living alone; $\beta = 0.05$, $t(2{,}614) = 2.49$, $p < 0.05$), being older than 55 ($\beta = 0.07$, $t(2{,}614) = 3.61$, $p < 0.05$), leadership role (ref. = no leadership role; $\beta = -0.04$, $t(2{,}614) = -2.11$, $p < 0.05$), and having sufficient protective equipment ($\beta = 0.19$, $t(2{,}614) = 9.66$, $p < 0.05$) are significant predictors (Adj. $R^2 = 0.045$). For mental exhaustion, living alone ($\beta = 0.05$, $t(2{,}614) = 2.83$, $p < 0.05$), being younger than 36 ($\beta = 0.06$, $t(2{,}614) = 2.80$, $p < 0.05$), and having sufficient protective equipment ($\beta = 0.23$, $t(2{,}614) = 11.97$, $p < 0.05$) are significant predictors (Adj. $R^2 = 0.056$). Finally, for general health, living alone ($\beta = -0.04$, $t(2{,}614) = -1.97$, $p < 0.05$) and having sufficient protective equipment ($\beta = -0.14$, $t(2{,}614) = -7.40$, $p < 0.05$) are significant predictors (Adj. $R^2 = 0.021$).

2.4. Discussion

In this brief research report we have investigated whether healthcare employees who work with COVID-19 patients report lower well-being and whether differences exist within that group.

Our results confirm that healthcare workers who treat COVID-19 patients experience more sleep problems and physical exhaustion compared to healthcare workers who do not treat COVID-19 patients. Furthermore, some personal and work characteristics present higher well-being risks.

In the light of the extant literature it should be acknowledged that in our study, effects are small to medium. In some of the other contexts, well-being appears to have decreased more drastically (Shreffler et al., 2020). What is more, mean scores still appear relatively acceptable (e.g., the lowest group score for general health is 7.24). This may point to the fact that the Netherlands has a relatively well-organized healthcare system (Daley et al., 2013). Nevertheless, our study contributes to the literature by, firstly, comparing effects across groups of healthcare workers, and secondly, presenting data from a non-Asian context.

There are a few limitations to discuss. First, whilst we employ validated scales, due to practical constraints in executing our survey we were not able to use clinical validated scales. Second, our data are collected in May-June 2020, right after the 'first peak', and the COVID-19 crisis as well as effects on well-being have developed since. Similarly, our cross-sectional design limits causal inference. Ergo, future research can improve on our current design by using validated tests, and employing longitudinal designs to track healthcare worker well-being over time.

Considering the practical implications of our study, we urge healthcare leaders, managers, and HR professionals to maintain healthcare worker well-being. Whilst a pandemic is hard to control, there are best practices on how to help healthcare workers deal with the consequences through, for example, job redesign, counseling, a behavioral health hotline, stress management webinars, respite rooms and creating celebratory rituals (Wei et al., 2020). Herein, our results show healthcare leaders should pay special attention to the groups of healthcare workers who appear disproportionally affected regarding either general health, physical or mental well-being, or sleep. Additionally, our results may fuel a number of questions to be discussed. For example, should more vulnerable healthcare workers (e.g., elderly female) be less actively deployed among COVID-19 patients? Which job resources help healthcare workers to deal with COVID-19 stressors including threats of infection, insecurity, work pressure, emotional demands, and work-family conflict (Foley et al., 2020; Kniffin et al., 2021)? How can healthcare workers be stimulated to share leadership to actively improve their own working conditions? The results also emphasize the grave importance of sufficient protective equipment. In conclusion, healthcare leaders are required to actively anticipate the evolution of this pandemic in order to maintain healthcare worker well-being; studies like these may help them to do just that (Torbay, 2020).

Chapter 3

Text mining and employee well-being

This chapter is based on the following published article:
Van Roekel, H., Wigger, E. F. J., Veldkamp, B. P., Bakker, A. B. (2023). What is work engagement? A text mining approach using employees' self-narratives. *Applied Psychology: An International Review.* DOI: 10.1111/apps.12501.

Abstract

We introduce text mining to study work engagement by using this method to classify employees' survey-based self-narratives into high or low work engagement and analyzing the text features that contribute to the classification. We used two samples, representing the 2020 and 2021 waves of an annual survey among healthcare employees. In the first study, we used exploratory sample 1 (n = 5,591) to explore which text features explain work engagement (unigrams, bigrams, psychological, or linguistic). In the second study, we confirmed whether features persisted over time between exploratory sample 1 and confirmatory sample 2 (n = 4,470). We find that psychological features classify employees across two samples with 60% accuracy. These features partly validate the literature: high-engaged employees refer more to affiliation and positive emotions, and low-engaged employees refer more to negative emotions and power. We extend the literature by studying linguistics: high-engaged employees use more first-person plural ('we') than low-engaged employees. Finally, some results question the literature, like the finding that low-engaged employees refer more to their managers. This study shows text mining can contribute by confirming, extending, or questioning the literature on work engagement and explores how future research could build on our findings with survey-based or *in vivo* applications.

3.1. Introduction

Whether employees are engaged in their work or not has important consequences for employees themselves, the organizations they work for, and the clients they work with. Engaged employees are full of energy, are dedicated toward work, and are often completely immersed in their work activities (Schaufeli and Bakker, 2010). They also experience more positive emotions, think in novel ways, and show better performance (Bakker et al., 2014; Christian et al., 2011). Since the emergence of the concept of work engagement, organizational scholars have been studying its presence, predictors, and outcomes (Bakker and Demerouti, 2017). However, most of the literature assesses work engagement with structured data, that is, measurement scales, and there have been few attempts to innovate measurements. Although structured data have allowed scholars to understand the phenomenon of work engagement, a drawback is the limited potential for new theoretical or applied discoveries (Balducci and Marinova, 2018).

At the same time, within organizations, a vast pool of data, in the form of unstructured (non-predefined) text data, remains scarcely studied. For example, employees generate and share large amounts of written text with each other. Those qualitative data may potentially offer new insights in work engagement and add to the more traditional structured approaches to data analysis. This is because unstructured data are not limited to predefined categories, present the multidimensionality of a phenomenon, and allow to compare these dimensions simultaneously (e.g., combining linguistic and substantive patterns) (Balducci and Marinova, 2018). Text mining offers a unique approach to unlock these insights as it is a method to analyze large amounts of text in a relatively short timeframe (Jurafsky and Martin, 2017). Its benefit compared with traditional quantitative or qualitative research is that it is able to analyze unstructured text, but on a large scale and replicable across studies. There are quite a few studies that show its potential in a variety of disciplines (e.g., Pang et al., 2020), but, although declared a future research avenue, few attempts have been made regarding organizational research (Kobayashi et al., 2021).

Therefore, the purpose of the current study is to explain work engagement through text mining methods by attempting to classify employees' survey-based self-narratives into high or low work engagement and analyze the text features that contribute to the classification. The research question that guided our study is as follows: to what extent can we explain work engagement by analyzing self-narratives through text mining? Using two samples, representing two waves of an annual survey among Dutch healthcare employees during 2 years of COVID-19, this paper conducts two studies to answer that question. We tested multiple text features: unigrams, bigrams, psychological features and linguistic features. For the psychological features, we conducted a preselection based on the Job Demands-Resources (JD-R) theory. Next, for the first study, we used exploratory sample 1 to explore which features explain work engagement. We then formulated hypotheses based on the main themes that emerged from the features. For the second study, we used both exploratory sample 1 and confirmatory sample 2 to analyze to what extent text features persist over time, across survey waves.

Our study contributes to the literature by being the first to explain work engagement by text mining self-narratives. Theoretically, we increase the understanding of work engagement as a concept. Some of our results confirm the duality of the JD-R model when we find low-engaged employees tend to mention job demands whereas high-engaged employees tend to mention job resources (e.g., Bakker et al., 2023; Bakker, 2022; Wang, Zhu et al., 2020). Yet because our analysis is exploratory, we are also able to extend and question extant findings. We find linguistic patterns may be markers of work engagement (e.g., Frankling and Thompson, 2005) and observe features that question the literature, like the finding that low-engaged employees mention their managers more often (Toegel et al., 2013). At the same time, we also discuss how our application of text mining is limited in terms of the accuracy with which we are able to explain work engagement, as well as how particular sample characteristics like age and gender may influence the results. Second, methodologically, our findings open multiple avenues for survey-based and *in vivo* applications of text mining. We discuss how text mining could support or complement structured forms of data collection (Kobayashi et al., 2021; Jurafsky and Martin, 2017), or be used to analyze existing unstructured data in organizations like emails or intranet posts (He et al., 2012). Finally, we explore how our study may have practical implications in the screening and identification of groups of employees based on work-related well-being challenges (e.g., He et al., 2012; Day et al., 2007).

3.2. Theoretical background

3.2.1. Defining, modeling and measuring work engagement

Work engagement is a work-related and positive state of mind, characterized by vigor, dedication and absorption. Vigor refers to a high level of energy and preparedness to invest effort in activities. Dedication refers to enthusiasm and strong involvement with one's work. Finally, absorption is a state of complete immersion in one's work (Bakker et al., 2014; Schaufeli et al. 2002). Whereas vigor and dedication are considered core dimensions of work engagement, absorption is considered an additional dimension

(Schaufeli et al., 2001). Whereas our knowledge of work engagement has increased in the past years, there are several remaining questions, for example, on its social-psychological origins and the effectiveness of work engagement interventions (Bakker, 2022; Knight et al., 2019).

Antecedents of work engagement are often studied within JD-R theory, a theory within organizational psychology that explains how job characteristics affect employees through a dual process (Bakker et al., 2023; Bakker and Demerouti, 2017; Schaufeli et al., 2009). In the health impairment process, demanding job characteristics—'aspects of the job that require sustained physical, emotional, or cognitive effort'—cause job strain (including burnout) and health complaints. Burnout refers to the state when employees experience chronic feelings of exhaustion and a cynical attitude towards work and the people with whom they work (Bakker et al., 2014). In the motivational process, resourceful job characteristics—'aspects of the job that help to either achieve work goals, reduce job demands (…), or stimulate personal growth'—foster motivational outcomes (including work engagement) and job performance (Bakker et al., 2014, p.392). In addition, job demands and resources are proposed to interact: resources may weaken the impact of demands on burnout, whereas challenge demands may strengthen the impact of resources on engagement (Bakker et al., 2014; Boyd et al., 2011).

Over the years, many resources that stimulate work engagement have been identified. Generally, they have been classified into one of two categories: situational and individual factors (Bakker and Demerouti, 2008). As explained above, antecedents of work engagement are mainly job resources. These include job characteristics like social support from colleagues, task significance, and autonomy as well as leadership-related factors like having a good relationship with supervisors and experiencing transformational leadership (Christian et al., 2011). In addition, individual factors have been found to explain work engagement. For example, employees with higher emotional stability, extraversion and conscientiousness are more likely to report higher work engagement. Besides these higher-order personality factors, lower-order factors—factors that are more malleable—have been found to predict work engagement, for example, self-efficacy and optimism (Mäkikangas et al., 2013). And employees who are more pro-active, tend to be more engaged and even positively influence their co-workers through practices of job crafting (Tims et al., 2015; Bakker et al., 2012).

In turn, studies have shown that work engagement can have far-reaching effects (Christian et al., 2007; Knight et al., 2017; Lesener et al., 2020). Research on work engagement shows positive relationships with more active positive emotions and more novel thinking (Bakker et al., 2014). What is more, there is abundant research that shows work engagement increases task performance (e.g., Christian et al., 2011), although studies also indicate that we know relatively little about the boundary conditions of these effects (Kane-Frieder et al., 2014).

The studies described above commonly measure work engagement with multidimensional scales. The most used scale that defines work engagement as the combination of vigor, dedication and absorption, is the Utrecht Work Engagement Scale (UWES; Schaufeli et al., 2006). Other measures of engagement are very similar to the UWES (see Schaufeli and Bakker, 2010; 2022), or measure concepts that are fundamentally different from engagement. For example, May et al. (2004) and Rich et al. (2010) developed the Job Engagement Scale, which includes cognitive, emotional, and physical engagement. According to Schaufeli and Bakker (2022, p. 285), 'the wording of the items shows a striking resemblance with those included in the absorption, dedication, and vigor subscales of the UWES, respectively'. The latter authors also discuss other instruments to assess engagement, including the instrument by Soane et al. (2012) and Shuck et al. (2017). Bakker et al. (2023, p.286) conclude that the items show considerable overlap with the vigor and absorption subscales of the UWES, whereas some of the alternative instruments that aim to assess engagement in fact assess affective organizational commitment and extra-role behaviors.

Although new scales have been developed since (Schaufeli et al., 2017), there have been few attempts to innovate measurement. For example, Bakker et al. (2014) point out that most of the research on work engagement has not attempted to link the concepts to observable outcomes. At the same time, there is some criticism on the UWES, including the fact that factor analyses have not always been able to distinguish between the three components of work engagement (Schaufeli and Bakker, 2022). Here, a new method like text mining may help optimize the measurement of work engagement by approaching it in a completely different way. Similarly, a particular bias of maintaining the same measurement methods is that these structured data limit the potential for new theoretical or applied discoveries (Balducci and Marinova, 2018). Text mining may allow new insights into what observable behaviors of employees are affected by work engagement. Below we address this issue further.

3.2.2. Considering unstructured data to measure work engagement

The vast majority of data in an organization are unstructured. Unstructured data refer to 'a single data unit in which the information offers a relatively concurrent representation of its multifaceted nature without predefined organization or numeric values' (Balducci and Marinova, 2018, p.558). For example, employees continuously exchange spoken or written text via conversations, email, or texting. These text data are seldom used in studies but may present new insights for the study of work engagement through three advantages over structured data. First, structured data (like survey scales) are always limited to the way they are defined and operationalized. In contrast, unstructured text data are neither predefined nor categorized, and this may lead to new insights. Second, unstructured data are multifaceted. There are multiple potential facets to unstructured data to be studied (e.g., there are linguistic and substantive properties to text). Third, unstructured data offer concurrent representation: through analyzing facets simultaneously (e.g., the combination of linguistic and substantive patterns), we can learn about different phenomena at the same time (Balducci and Marinova, 2018).

Although unstructured data offer new opportunities for research, structured data are important too. Unstructured data provide, besides its general format being either text-based or image-based, little certainty. This type of data does not allow for easy sorting, searching, analyzing, summarizing, or visualizing. Structured data, on the other hand, provide certainty in measurement and analysis as the data are set in predetermined categories or values. This type of data is easily stored, searched, analyzed, summarized, and visualized. The data contain exactly what could be expected and allow for accessible, unbiased analysis. It has been an important part of theory-based research as theory is operationalized into a specific measurable form. In comparison, unstructured data require a different, more thorough approach to hold value (Hanig et al., 2010).

Although unstructured data have been underused with regard to work engagement, previous studies have shown that free-form text can be a rich source of data that contains important insights about mental wellbeing and allows identification and screening for mental diseases. For example, research has shown that the content of the speech of schizophrenics differs substantively from non-schizophrenics (Franklin and Thompson, 2005; Rosenberg and Tucker, 1979). There are some specific examples of text mining in psychology and organizational research. Pang et al. (2020) succeeded in predicting 24-character strengths, like gratitude, zest and leadership, based on Twitter language. This study indicated that one can use text mining to measure the character strengths of large populations. Similarly, La Bella et al. (2018) used text mining to track perceived organizational leadership styles almost real-time with Twitter messages. Examples in clinical settings include the screening of posttraumatic stress disorder in self-narratives (He et al., 2012) and the identification of trauma patients (Day et al., 2007). When employing text mining for work engagement, we hope to explore whether and how employees high in work engagement may display different features from employees low in work engagement.

One reason for the limited attention to analyzing textual data may be that traditionally analyzing text was a time-consuming endeavor as manual coding was the only option. However, new techniques derived from machine learning and statistics may enable to study work engagement and other concepts with unstructured data (Kobayashi et al., 2021). One particularly promising avenue to innovate is by text mining as it is a general methodological framework to analyze large corpora of text (Jurafsky and Martin, 2017). Hence, text mining offers large-scale text analysis in short timeframes, only bottlenecked by computing power and the fact that often large amounts of texts are required to generate insights. Its benefit, therefore, compared with traditional quantitative or qualitative research is that it is able to analyze unstructured text, but on a large scale and replicable across studies. Only recently organizational scholars have suggested that this approach could be used to assess organizational concepts like burnout (Kobayashi et al., 2021).

Text mining refers to the analytical process that aims to generate insights or test hypotheses using unstructured text data (Kao and Poteet, 2007). The data are systematically collected, cleaned and transformed (a process referred to as preprocessing), after which one of multiple text mining operations can be applied to generate insights from the text data. Texts can be analyzed based on textual patterns and linguistic features, as well as dictionary approaches, considering the words used in the texts. Finally, postprocessing requires the interpretation and evaluation of the results by applying specific domain knowledge to them and validating the data (Kobayashi et al., 2018).

There are many text features that can be analyzed through text mining. First, textual patterns such as bag-of-words approaches look at the occurrence of words in texts and try to understand the corpus based on word counts (Aggarwal and Zhai, 2012). Similar approaches include the use of n-grams, which refers to word combinations of two (e.g., ‘working day’), three (e.g., ‘busy working day’), or more words. The advantage of using bag-of-words and n-grams lies in their simplicity. The only information lost is the position of the words or n-grams in the text, which means it is a true-to-source feature to analyze. The downside, however, is that respondents with different background characteristics like educational background or social status may use different words to convey identical information. Patterns might emerge based on characteristics that are unrelated to the research at hand.

Second, there are dictionary approaches, which tag words in the texts with categories the words belong to. In our analysis, we can use these categories to understand the texts. An example of such a dictionary approach is Linguistic Inquiry and Word Count

(LIWC). LIWC counts words in linguistic and psychologically meaningful categories (Tausczik and Pennebaker, 2010). Dictionary approaches resolve the issue of textual pattern features as the underlying meanings and categories of the words are analyzed, rather than the words themselves. However, the downside is a loss of information as the words themselves are not analyzed further. In choosing the features, one could either test all available features or apply some sort of *a priori* selection. *A priori* selection is often applied to prevent overfitting. Overfitting is a common problem in machine learning where a model performs well on the training data but fails to generalize to new, unseen data. This happens because the model is trained on a specific set of data that might not contain a fully balanced representation of all words that do or do not contribute to the classification of the variable. The model will fit to the training data as specifically as possible, even though there might be false patterns that do not hold over multiple samples. Overfitting can be avoided by using a larger and more diverse dataset for training, as well as using regularization techniques to prevent the model from learning overly complex patterns in the data. Our text mining approach was both theory- and data driven. Specifically, we used JD-R theory to select psychological features based upon their resemblance to any aspect of the definition, dimension or items of work engagement. In sum, in the present study we hope to gain new insights into the concept of work engagement using text mining. Across two studies, we will compare bag-of-words, bigrams, and LIWC dictionary approaches (psychological process and linguistic features) to explore the possibilities offered by text mining.

3.3. Methods

3.3.1. Procedure, samples and data

We used data from two samples, representing two waves of an annual survey among Dutch healthcare employees who are members of *Stichting IZZ,* a collective of healthcare employees and employers in the Netherlands[1]. This foundation has over 400,000 members of which around 210,000 are healthcare employees, who make up a notable share of the population of around 1.7 million healthcare employees in the country (CBS, 2022a; Van der Fels, 2020). The annual survey, executed since 2018, is used to monitor how healthcare employees perceive their work and well-being. It presents an opportunity for employees to share their experiences, which are then shared (at the group level) with healthcare organizations, governments, societal partners and media. The survey has, among else, been helpful in informing these parties about the

1 The data collection for this study was reviewed and approved by the Faculty Ethical Review Committee of Utrecht University (nr. 2019-004).

challenges COVID-19 posed for healthcare employees. Besides, to add an extrinsic motivation to finish the survey, participants could choose to participate in a separately organized giveaway (with products that stimulate well-being).

For the first sample, data were collected in May and June 2020. For the second sample, data were collected in May and June 2021. Table 1 shows how we arrived at our final sample. All members of the collective that provided an email address were sent an invitation to participate in the survey via email. For a response to be valid, respondents had to provide informed consent and indicate they were currently working in healthcare. At the informed consent page, respondents were informed about the goal of the survey, procedures of participation and opting out, data storage and usage, and the possibility to get in touch with the researchers. The survey itself consisted out of multiple open and closed questions on employee well-being in healthcare, including the questions presented in this study. We did not employ attention checks. All questions used in this study, except the text mining question, used forced response. The text mining question was placed at the end of the survey as it would take considerable time. Therefore, as Table 1 indicates, respondents with partial responses corresponded with respondents who did not fill in the text mining question—these were removed from the dataset. Besides, to be included in our final sample, we set the minimum number of written words at 20. Finally, as in our study we will compare employees in the top 10% with those in the bottom 10% of work engagement scores, Table 1 also presents this subsample. The methodological choices made above will be elaborated on below.

Table 1 *Sample justification*

	Sample 1 (2020)	Sample 2 (2021)
No. of survey invitations	138,382	133,322
No. of responses recorded	19,772	8,955
~ Response rate (recorded responses / invitations)	14.29%	6.72%
No. of valid responses*	12,630	8,132
No. of responses to text mining question	5,976	5,016
~ % of partial responses (no answer text mining / valid responses)	52.68%	38.32%
No. of responses with 20 words or more (final sample)	5,591	4,470
~ Response rate for final sample (final sample / invitations)	4.04%	3.35%
No. of responses in top / bottom 10%** (subsample)	1,119	894

Note. *Respondents who gave consent to participate, and were currently working in healthcare. **Number of responses in the top or bottom 10% of work engagement scores.

Table 2 presents the characteristics of the respondents in the final sample. The respondents are representative for the population of Dutch healthcare employees in terms of gender (84.3% of employees are female) but somewhat less representative in terms of age (employees in the population are younger: 34% are younger than 35 and 24.2% are older than 55) (CBS, 2020a). Especially the gender composition, with a vast majority of female employees, is a typical (but not unique: e.g., in Dutch primary education, 87% of teachers are female; Ministerie van OCW, 2021) characteristic of healthcare sectors. While the majority in the Netherlands is very large, a WHO-report shows across the world women constitute around 70% of the healthcare workforce (2019). What is more, the same report indicates that the small minority of men in healthcare is more likely to hold leadership positions. This leadership gap has systemic roots in gender roles (see e.g., Ryan et al., 2016): men and women in healthcare (are expected to) work in different jobs. We should take into account that such factors could affect work engagement. Besides, although our sample is older than the population of healthcare employees, the population of healthcare employees is ageing rapidly (the group of healthcare employees aged 55 and over saw a 9% increase in just 10 years; Van Wijk, 2020). Finally, nursing/home care, hospitals and disabled care constitute the biggest healthcare branches within the Netherlands, and are also the largest in our sample. However, although hospitals are the biggest group in our sample, within the population nursing/home care is bigger (with a total of 28% of healthcare employees in the population; Van Wijk, 2020). In sum, our samples of healthcare employees are fairly representative for healthcare.

It is important to note that there may be systemic reasons to expect differences in work engagement based on demographic characteristics. Table 3 shows that to some extent work engagement varies across age, gender and healthcare branch. Most notably, work engagement is generally higher among women, among 46-55-year old employees (not taking into account the youngest and oldest categories, which both have low *n)*, and among employees in nursing / home care (Appendix 6 elaborates on these differences). This may have consequences for our study's external generalizability. This study analyzes text-based features among a fairly representative sample of healthcare employees, but the particular sample characteristics (e.g., distribution of age and gender) and how work engagement relates to these characteristics may limit generalization to different sectors.

Table 2 *Sample characteristics (n_1 = 5,591 and n_2 = 4,470)*

	Sample 1	**Sample 2**
Gender		
Female	4,731 (84.6%)	3,816 (85.4%)
Male	845 (15.1%)	636 (14.2%)
Rather not say	15 (.3%)	18 (.4%)
Age		
-25	48 (.9%)	36 (.8%)
26-35	447 (8%)	240 (5.4%)
36-45	991 (17.7%)	709 (15.9%)
46-55	1,711 (30.6%)	1,366 (30.6%)
56-65	2,327 (41.6%)	2,081 (46.6%)
66-	63 (1.1%)	36 (.8%)
Unknown	4 (.1%)	2 (< .1%)
Healthcare branch		
Hospitals	1,983 (35.5%)	1,582 (35.4%)
Nursing / Home care	1,343 (24%)	1,175 (26.3%)
Mental healthcare	911 (16.3%)	681 (15.2%)
Disabled care	981 (17.5%)	743 (16.6%)
Other	373 (6.7%)	289 (6.5%)

Finally, to protect the privacy-sensitive information that participants provided in their self-narratives, data are stored on secure servers in compliance with privacy regulations and not made publicly available. We do present appendices 1-7 that provides extra information on the research process: an overview of the *included features,* the *R script* for our analyses, our approach to deciding the *cutoff,* an overview of all *significant features,* an overview of the relative *feature importance* to the models, *additional analysis* on the role of demographics, and an additional analysis using a different classifier (*Naive Bayes).*

Table 3 *Work engagement across sample characteristics*

Work engagement score		
	Sample 1	**Sample 2**
Gender		
Female	3.89 (.62)	3.81 (.65)
Male	3.78 (.67)	3.70 (.72)
Age		
-25	4.01 (.54)	3.68 (.59)
26-35	3.81 (.56)	3.70 (.62)
36-45	3.84 (.59)	3.79 (.61)
46-55	3.90 (.62)	3.83 (.66)
56-65	3.86 (.66)	3.77 (.69)
66-	3.98 (.66)	4.16 (.56)
Healthcare branch		
Hospitals	3.84 (.62)	3.75 (.68)
Nursing / Home care	4.00 (.62)	3.91 (.66)
Mental healthcare	3.76 (.62)	3.73 (.63)
Disabled care	3.84 (.64)	3.77 (.66)
Other	3.92 (.62)	3.75 (.67)

Note. This table presents means and standard deviations. Appendix 6 presents statistical tests for work engagement scores across sample characteristics and additional descriptives.

3.3.2. Work engagement scale

The UWES-9 work engagement scale includes nine items on three dimensions: vigor, dedication and absorption (Schaufeli et al., 2006). All dimensions were measured with three items on a 5-point Likert scale ranging from 'Never' (1) to 'Always (daily)' (5). Example items are: 'At my work, I feel bursting with energy' (vigor), 'I am proud of the work that I do' (dedication), and 'I feel happy when I am working intensely' (absorption). The items were summed to create an overall index of work engagement. The reliability of the overall scale was good, Cronbach's alpha was .908 for sample 1 and .910 for sample 2.

3.3.3. Self-narrative question

Below we present the English translation of the question that was shown to respondents to write their self-narrative:

We have one additional question about how you have experienced your work during this time of COVID-19. We would like to take a closer look at your personal experiences. Could you summarize what you have experienced? How have you experienced the past few months? What impact has this had? How are you feeling now, physically and emotionally? How do you view your work now? And how do you look forward to the coming months?

This multifaceted question functioned as a writing prompt to guide the content of the self-narratives (Kroll and Reid, 1994). Writing prompts make writing easier when they, in our case, promote the structure of the story that respondents are expected to write (Hudson et al., 2005). The question was drafted purposefully and in reiterative discussion between all the authors of this study and contained multiple subquestions that each served their own purpose. The first subquestion was general (Could you summarize what you have experienced?), after which the second subquestion specified the time period (How have you experienced the past few months?). Third, we asked about the consequences of these experiences (What impact has this had?). Fourth, we referred to the energy continuum of exhaustion versus vigor (How are you feeling now, physically and emotionally?). Fifth, we referred to the identification continuum of dedication versus cynicism (How do you view your work now?). Finally, we asked about participants' future perspective (And how do you look forward to the coming months?). After this, a text box provided participants ample opportunity to share their self-narratives.

3.3.4. Analysis

The process of text mining involves four basic steps: (1) data preprocessing; (2) training on a subset of the data; (3) testing on a different subset of the data; and (4) interpreting the results. We will explain these steps below.

Data preparation

In the first data preparation step, the corpus of all texts was cleaned, and features were extracted and selected to prepare for data analysis (Wickham et al., 2022a; 2022b; 2022c; Silge and Robinson, 2016). Packages and code used can be found in Appendix 2. We checked whether we needed to apply criteria for minimum or maximum number of words in the self-narratives. An explorative analysis of the data indicated that respondents who replied with fewer than 20 words in their self-narratives commonly responded with variations of 'I do not have anything to share'. This is not a substantive answer to the question prompt—yet it occurred many times in the initial dataset. We decided to require respondents to have written 20 words at minimum to avoid meaningless self-narratives, but we did not apply a maximum as no self-narrative appeared extremely long. The second part of the data preparation aimed to tokenize the text by removing punctuation, numbers, and capitalization, and by splitting the texts word by word (Benoit et al., 2018). The resulting list of words was spell-checked by one of the researchers for all words that occurred at least a total of 10 times to find spelling errors or gibberish that would be included in the model. One returning issue concerned the occurrence of abbreviations alongside the same abbreviations written out in full. This included both general abbreviations as well as job-specific abbreviations. As abbreviations were often unclear and seemed to vary on a text-by-text basis, these abbreviations were not manipulated to full terms. No other issues were found based on this quality check of the data, and after this quality check we proceeded with the data. Further cleaning steps were the removal of frequently occurring stop words (e.g., 'the', 'a') that do not discriminate texts or add little to no meaning to texts. The list of stopwords filtered is based on the Dutch stopwords list as compiled by the Snowball stemming project, this list is included in the corpus package (Perry, 2017; Porter, 2001). Finally, stemming the tokens, reducing all words to their stem, was done to ensure various inflections of the same word are counted together (e.g., 'working', 'worked' and 'work' become 'work'). For sample 1, this step reduced the total number of words by 48.44% (from 632,174 to 325,963) and the unique number of words by 29.41% (from 22,524 to 15,899). For sample 2, this step reduced the total number of words by 48.01% (from 443,668 to 230,653) and the unique number of words by 17.77% (from 15,573 to 12,806).

The other steps in data preparation were feature extraction and feature selection. This study used features based on bag-of-words approaches and dictionary approaches to explain work engagement. Importantly, below we describe how the features were selected for sample 1. For sample 2, we only used the features that contributed to the classification into high and low work engagement in sample 1.

First, the bag-of-words approach assumes no relationship between the order of the words in a text and the meaning of the text. The words are as they are, independent from other words in the text. This approach was used for generating unigrams (one word, e.g., 'happy') and bigrams (two words, e.g., 'not happy'). Second, the dictionary approach used the LIWC dictionary to tag words with categories belonging to Psychological Processes or Linguistic Dimensions (Pennebaker et al., 2015). The translated Dutch version of LIWC 2015 has 67 categories for which the words can be matched (Van Wissen and Boot, 2017). The features selected were both theory-driven and data-driven. For the Linguistic Dimensions, we explored all features available. For the psychological features, we selected features using the JD-R theory: we checked whether they reflected an aspect of the definition of work engagement, its dimensions, or items. We selected features from the affective, social, perceptual, and biological processes, drives, time orientations, relativity, and personal concerns (Pennebaker et al., 2015). Although this means that there was some *a priori* selection of features, the selection was necessary to limit the scope of the study. Additionally, because of the theory-related nature of the LIWC psychological process features, *a priori* selection based on theory could prevent overfitting. Through *a priori* selection of psychological process features, we narrowed the amount of psychological process features from 54 categories (including all overarching categories and more specific subcategories) to 25 categories. This was done to remove non-work-related features, as these would be less relevant for our purposes of explaining work engagement. That is, in this specific analysis we were looking for factors related to work, not for other factors. That is not to say that work engagement is irrelevant for employees' private lives, as research suggests otherwise (e.g., Wood et al., 2020). It only means that we slightly narrowed the scope of our analysis. Comparing the features to the definition, dimensions or items of work engagement provided a good measure for selection. For example, the category 'Time orientations' (including the categories 'past focus', 'present focus', and 'future focus') was included as work engagement is related to time orientations. The dimension absorption includes the phrase 'whereby time passes quickly' (Bakker et al., 2014, p. 391). In contrast, some subcategories of 'Personal concerns', like the 'Leisure' category with words like 'home', 'chat' and 'movie' were deemed less relevant for our endeavors (Pennebaker et al., 2015). Additionally, to prevent overfitting for the Random forest model and to ensure robust features were kept, feature selection was applied for each of the feature representations using chi-squared tests (Forman, 2004; 2003). Features were kept based on the criteria of significant chi-squared outcome for features that occur at least 10 times in the first sample dataset (He et al., 2012). Appendix 1 presents an overview of all included features and specifies why features were included.

Training and Testing

The training phase consisted of learning from a first subset of the sample to understand how the text is related to the outcome variables. In the exploratory phase, our first study, we tested a number of settings and chose those in which the models performed better for sample 1. First, we used the full UWES-9 scale for the classification of employees (Schaufeli et al., 2006). We did explore whether it would make sense to focus on the energetic component of work engagement by using only one of the dimensions of work engagement (vigor)—or a combination of vigor with exhaustion (a dimension of burnout, Schaufeli et al., 1996)—to classify employees. The argument here was that perhaps this would lead to better classification as studies indicate the energetic component of work engagement is more sensitive than the other components (Bakker and Demerouti, 2017). We decided to focus on the UWES-9 as most studies of work engagement use this full scale. Second, as Forman mentions in his critique on text classification feature selection methods, classifying minority classes is a pitfall for feature selection methods that use scoring methods based on outcome variables (2004). We therefore carefully decided on the criteria for classifying employees into high or low work engagement (Appendix 3 explains this process in more detail). We explored relative (percentages) and absolute groups (e.g., scores below 2 and above 4 on a 5-point scale). We found the sample is unbalanced: more healthcare employees tend to be relatively high-engaged. Therefore, we chose to use a relative, 10% cutoff. For the confirmatory phase, our second study, the same cutoff was set *a priori*.

Hence, to correctly classify whether an employee is high work engaged or not, we decided to select employees with a self-narrative of at least 20 words, who had a work engagement score in the highest 10% versus a work engagement score in the lowest 10% of each sample. For the first study, sample 1 (n = 1,119) was split into a training set containing 80% of the self-narratives and a testing set containing the remaining 20% of the self-narratives. For the second study, we wanted to analyze how the features persist over time and across survey waves, so we used sample 1 in its entirety as the training dataset and sample 2 as the testing set. Table 4 reiterates the way the two studies were set up.

Table 4 *Samples and studies*

	Study 1	**Study 2**
Type of study	Exploratory	Confirmatory (with hypotheses)
Goal	Explore which text features explain work engagement in sample 1	Test the extent to which the text features persist over time (between sample 1 and 2)
Used training set	80% of sample 1 (n = 895)	Sample 1 (n = 1,119)
Used testing set	20% of sample 1 (n = 224)	Sample 2 (n = 894)

The purpose of splitting the sample into a training and testing set, is to learn to recognize high versus low work engagement in the training set using the mentioned text features, after which the testing set can be used to assess its performance in recognizing high and low work engagement for previously unseen data. We used Random forest, a machine learning model that generates many decision trees trained and tested using resampling of the sample data. Each tree randomly samples a predefined number of features per split and decides based on the feature that best distinguishes the classes at that point in the tree. After all trees are built, Random forest calculates the best scoring features based on the classification scores per tree with and without the feature. If the trees classify worse when the feature is excluded from the tree, that means the feature has some explanatory power (Kotsiantis et al., 2007; Breiman, 2001).

Random forest allows for hyperparameter optimization, which is the process of modifying the model settings for better model performance. The Random forest model was applied using the randomForest package in R (Liaw and Wiener, 2002). The randomForest package allows for different settings for the nodesize, mtry, and ntree hyperparameters. For study 2, the Random forest hyperparameter optimization was done using the caret package by doing grid search for the minimum node size (nodesize) and the number of variables to sample as candidates for each split of a node (mtry) (Kuhn, 2008). A smaller nodesize hyperparameter value allows for more splits in the tree, resulting in a more complex tree (Breiman, 2001). Additionally, the number of trees (ntree) was optimized by building the Random forest model with 100, 250, 500, and 1000 trees. No noticeable improvement was found after 250 trees for any of the models, test set error rate was lowest with 250 trees, and highest with 500 and 1000 trees, whereas OOB error rate only marginally improved.

For comparison, in study, 2 we also used a different classifier, Naive Bayes (Benoit et al., 2018; Lewis, 1998). Naive Bayes is a probabilistic machine learning algorithm that assumes features are independent of each other. The algorithm is well suited for high dimensional datasets such as text data because it is efficient due to scaling linearly with the number of predictors and datapoints. Despite the assumption of independence often being violated, Naive Bayes tends to deliver robust and accurate classification. Naive Bayes, Random forest, and other approaches such as support vector machines and logistic regression are commonly used in text mining classification problems. In data science there is no consensus on the best method as it depends on the data at hand. This means that in practice researchers use a variety of algorithms based on the conditions for the used case and pick the best performing algorithm.

The model results are evaluated primarily using a confusion matrix, the accuracy score and *p*-value for accuracy score compared with the no-information-rate (NIR), which is an accuracy value that always predicts the most frequently occurring class in the dataset (Kuhn, 2008). Appendix 2 presents the R code for our main analyses.

Interpreting

In the fourth step, we evaluated the text mining results by comparing them to the domain knowledge on work engagement. For study 1, we used an exploratory approach to interpretation, by comparing the exploratory results to the existing literature on work engagement after conducting the analysis. We developed several hypotheses that explicated the main themes emerging from the analysis of study 1. We use both our observations (our data from study 1) and potential explanations in the existing theories in the literature to inform our hypotheses. For study 2, we used a confirmatory approach to interpretation, by assessing the hypotheses formulated in study 1. These hypotheses guided our discussion in study 2, and enabled to assess whether the same features contribute to explaining work engagement over time (compare our approach to studies on scale development who also distinguish an exploratory and confirmatory phase, e.g., House et al., 2004).

3.4. Results study 1

In this section, we present our results in four steps. First, we described the groups of high and low-engaged employees in the sample. Second, we counted all the features that we analyzed in the self-narratives of these employees, and we assessed whether features are significantly more frequently observed among high or low-engaged employees. Third, we tested whether the features can be used in a Random forest model that can correctly classify employees into high or low work engagement (using the training and testing approach, as described in our methods). Fourth, we presented the features that contribute most to the accuracy of the model, as these features indicate best how self-narratives of high and low-engaged employees differ.

First, there are 5,591 respondents who answered the text mining question with 20 words or more. The mean number of characters in the self-narratives was 667.24 (*SD* = 442.01), and the mean number of words was 113.31 (*SD* = 76.21). The mean work engagement score was 3.87 (*SD* = .63). Table 5 presents the respondents that are in the highest and lowest 10% of work engagement scores. The variance within the lowest 10% is notably larger than the variance in the highest 10%. This shows that our 'lowest 10%' group is a varied of group employees ranging from very low on work engagement to moderately engaged.

Table 5 *Highest versus lowest 10% scores on work engagement*

	No. of self-narratives	**Mean score**	**Median score**	**Min. score**	**Max. score**
Highest 10%	559	4.87	4.89	4.67	5
Lowest 10%	560	2.60	2.67	1	3

Second, we counted the features in the groups of employees. Appendix 4 presents all features that are observed significantly more frequently in self-narratives of either high or low-engaged employees. When we present the most contributing features in Table 7, we use the information from this step to indicate which feature is observed significantly more among high or low-engaged employees.

Third, we employed Random forest to test whether these features can be used in a model to classify employees in the highest 10% or lowest 10% of work engagement. Table 6 presents the results of four models: unigrams, bigrams, psychological features and linguistic features. Table 6 first indicates how many features contributed to the models. Next, it shows how the models performed. We find that unigrams score best with a 62% accuracy score, whereas the other models have lower accuracy scores (for bigrams, 58%; for psychological features, 60%; and for linguistic features, 59%). Considering that a model based on randomization would have an accuracy score of 50% (an equal chance of true or false classification), we find that all models classify into high or low work engagement better than random. The model with the unigrams appears the most successful.

Table 6 *Results Random forest for unigrams, bigrams, psychological and linguistic features*

		Model 1: unigrams		**Model 2: bigrams**		**Model 3: psychological features**		**Model 4: linguistic features**	
No. of features		156		24		16		6	
Confusion matrix test set (actual/predicted)									
TP	FP	57	35	96	84	64	53	73	44
FN	TN	50	82	11	33	37	70	47	60
Model statistics									
OOB		32.63%		42.01%		42.23%		47.60%	
Accuracy		62.05%		57.59%		59.82%		59.38%	
NIR		52.23%		52.23%		52.23%		52.23%	
p-value (Acc > NIR)		0.002		0.062		0.013		0.019	

Note. The confusion matrix presents the numbers for: respondents that score low on work engagement which our model got right (TP), respondents that score low on work engagement which our model got wrong (FN, a type 2 error), respondents that score high on work engagement which our model got wrong (FP, a type 1 error), and respondents that score high on work engagement which our model got right (TN).

Fourth, we analyzed the features that best classify into high or low work engagement in the different models. For that, we assessed the discriminatory value of the features based on the mean decrease in accuracy of the models when a specific feature is excluded. Appendix 5 presents the importance of all the features in the models. Table 7 presents the (most) strongly contributing features. For the unigrams and bigrams, we translated the features from Dutch into English and provided them with a common stem to ease interpretation. For the psychological and linguistic features, the feature categories are presented, and if applicable the overarching category is presented between parentheses. Appendix 1 presents more information on the content of these categories (so does Pennebaker et al., 2015).

Table 7 should be interpreted as follows: the bigram 'goes well' contributes most to the accuracy of the bigrams model, and this bigram is significantly more present among employees with high work engagement. In contrast, the bigram 'from house', which is the second most contributing feature, is significantly more counted among employees with low work engagement. Likewise, we find that the psychological feature 'positive emotion', an LIWC dictionary with words like 'safe', 'trust', and 'beloved', contributes most to the accuracy of the psychological features model, and is significantly more counted among high-engaged employees. In contrast, the feature 'anger', a subdictionary of negative emotions, with words like 'aggression', 'stupid' and 'fight' and the second most contributing to this model, is significantly more present among employees low in work engagement. In our discussion, we interpret what themes are presented in the features and how this relates to the extant literature.

Table 7 *Contribution of features to the models*

	Model 1: unigrams*	**Model 2: bigrams**	**Model 3: psychological features**	**Model 4: linguistic features**
No. of contributing features				
	111	19	9	2
Most strongly contributing features				
1	few/little	**goes well**	**Positive emotion**	**3rd person plural**
2	specific	from house	Anger (Negative emotion)	**1st person plural**
3	**nursery home**	**past month**	Power (Drives)	-
4	**family**	high work pressure	**Social processes**	-
5	**good**	unsafe feeling	Present focus (Time orient.)	-
6	workplace	**allowed come**	Negative emotions	-
7	**suspected**	hours per	**Affiliation (Drives)**	-
8	**see**	meter distance	**Reward (Drives)**	-
9	high	**we go**	Work (Personal concerns)	-
10	**allowed**	usual work	-	-
11	**listening**	**colleague s**	-	-
12	face-to-face	**we good**	-	-
13	crisis	come work	-	-
14	**super**	**work we**	-	-
15	**caregiver**	**now then**	-	-
16	burdened	colleague does	-	-
17	**cohort**	**very good**	-	-
18	pension	direct contact	-	-
19	information	**our resident**	-	-
20	**talking**	-	-	-

Note. Features that are present significantly more among high-engaged employees are in bold, features present significantly more among low-engaged employees are in normal font. *The features presented in this column are the 20 most strongly contributing features and have a chi^2 -value of 3.7 or higher.

3.5. Discussion of study 1

For study 1, we used an explorative approach to assess whether we can explain work engagement through text mining. We found that models with unigrams, bigrams, psychological features and linguistic features can correctly classify healthcare employees into high or low work engagement with an accuracy of up to 62% (for the unigrams). Whether we will find similar results in the next study depends on two aspects of the features. First, it will depend on the translatability of the type of feature. Herein, we may expect differences between the types of features we use. A methodological explanation for the success of unigrams in this study is that the high number of features may allow for more discrimination between the groups. However, for study 2, the question is whether unigrams will translate over samples. Potentially, psychological and linguistic features will explain better across samples than unigrams or bigrams as they use dictionary approaches that measure underlying meanings and categories of words rather than the specific words themselves (Tausczik and Pennebaker, 2010).

Second, whether we find the same result in study 2 will depend on the translatability of the content of the feature. If we assume that healthcare employees' work engagement can be explained by factors that are time-insensitive, our models should perform similarly. We will therefore explore whether we observe a few grand themes among the features that contribute to our models. As the process of feature selection was partially data driven, there are many features and not all are readily interpretable or categorizable. However, across the models, three prominent feature themes emerge from the self-narratives, which we named: emotions (24 features), crisis (35 features), and affiliation (19 features) (Appendix 5 presents the coding). Below we introduce these themes, compare them with the extant literature and formulate hypotheses for study 2.

First, the models include strongly contributing features that address the positive or negative emotions in the self-narratives. For the unigrams, we find positive emotion words to be related to high-engaged employees (e.g., good) and negative emotion words to be related to low-engaged employees (e.g., burdened). For the bigrams, we find positive emotion word combinations to be related to high-engaged employees (e.g., goes well, very good) and negative emotion word combinations to be related to low-engaged employees (e.g., unsafe feeling). For the psychological features, we find positive emotions to be related to high-engaged employees, and anger and negative emotions to be related to low-engaged employees. These findings support the conceptualization of work engagement as a 'positive motivational state' (Bakker et al., 2014, p. 389) as well as the finding that positive emotions are positively related to work engagement (Ouweneel et al., 2012). Additionally, employees who experience job strain are less able to regulate their emotions (Bakker and Costa, 2014), and scholars have suggested that emotional instability may be a personal demand that affects work engagement (Lorente Prieto et al., 2008). Hence, our first hypothesis is as follows:

H1: *Referring to positive emotions contributes to explaining high work engagement, whereas referring to negative emotions contributes to explaining low work engagement.*

Second, the models include features that refer to the crisis during which the study was conducted, the COVID-19 pandemic. For the unigrams and bigrams, we find references to the crisis related to low-engaged employees (e.g., high work pressure, meter distance, usual work, direct contact; face-to-face [contact], crisis) and references to the absence of the crisis to be related to high-engaged employees (e.g., family, allowed; goes well, allowed [to] come). A sidenote here is that we indicated we have reason to expect that unigrams and bigrams translate less well across samples. Nevertheless, the features echo an emerging stream of literature that shows the pandemic may in many cases have deteriorated work engagement (Kniffin et al., 2021) and other aspects of well-being (Van Roekel et al., 2021; Wang et al., 2021). For many healthcare employees, COVID-19 caused higher stress levels and other negative health outcomes (e.g., Shreffler et al., 2020). There is an emerging literature on the effects of a crisis in the context of JD-R theory. Demerouti and Bakker (2023) argue the COVID-19 crisis has increased job demands. Besides, a crisis also tends to make resources scarce. They propose that during a crisis, employees who experience manageable job demands (and high job resources) will maintain higher engagement than employees who experience high job demands (and low job resources). At the same time, changes in engagement are likely not only caused by individual demands and resources but by a more complex interplay of individual and higher-level factors. In sum, our second hypothesis is as follows:

H2: *Referring to a crisis contributes to explaining low work engagement, whereas referring to a normal work context contributes to explaining high work engagement.*

Third, the models include features that refer to affiliation and social connection, which appears to explain high work engagement. For the unigrams, we find references to social contact (e.g., listening and talking) to be related to high-engaged employees. For the bigrams, we find multiple plural references to be related to high-engaged employees (e.g., we go, work we, and our resident). For the psychological features, we find social processes (e.g., 'talk' and 'love') and affiliation (e.g., 'friend' and 'social') to be related to high-engaged employees. And for the linguistic features, using 1st and 3 rd person plural is positively related to high-engaged employees (in addition, Appendix 4 show that low-engaged employees use significantly more singular forms but these do not contribute to the models). This supports studies that show the importance of affiliation for work engagement. First, experiencing social support is an important predictor of work engagement, and in turn, employees who are engaged offer more social support (Freeney and Fellenz, 2013). Another study explained how especially new employees' work engagement is highly affected by socialization in the organization (Saks and Gruman, 2018). Additionally, scholars have studied the concept of team work engagement, which indicates that work engagement is not merely an individual process but also part of a team process (Costa et al., 2014). Work engagement is contagious, and employees can collectively experience high levels of work engagement (Bakker, 2022; Bakker et al., 2016). Therefore:

H3: *Referring to affiliation contributes to explaining high work engagement.*

Having defined our hypotheses, we submitted a preregistration at the Open Science Framework that described the hypotheses as well as the plan of analysis. For study 2, our primary aim is to select the features of study 1, and to assess whether these features contribute strongly to the models in study 2. We will evaluate the success of these features both in terms of specific features as well as the themes that they represent (as referred to in the hypotheses). The next section describes the results of study 2.

3.6. Results study 2

In study 2, we repeated the analysis with both samples. Again, we present the results in four steps: we described the employees in the sample, we counted all features and tested whether features are significantly more frequently observed among high or low-engaged employees, we built the Random forest model, and we presented the features that contribute most.

First, we already introduced the first sample above. In the second sample that we add in this analysis, a total of 4,470 respondents answered the text mining question (20 words or more). The mean number of characters in the self-narratives was 583.89 (SD = 397.33) and the mean number of words 98.03 (SD = 67.99). The mean work engagement score was 3.79 (SD = .66). Again, we selected the respondents that are in the highest and lowest 10% of work engagement scores (Table 8). Like the first sample, the variance within the lowest 10% is notably larger than the variance in the highest 10%. Average scores also appear to be slightly lower for both groups compared with the first sample.

Table 8 *Highest versus lowest 10% scores on work engagement*

	No. of self-narratives	Mean score	Median score	Min. score	Max. score
Highest 10%	447	4.84	4.78	4.67	5
Lowest 10%	447	2.46	2.56	1.11	2.89

Second, we counted the features. Appendix 4 presents all features that are observed significantly *more* in self-narratives of either high or low-engaged employees. When we present the most contributing features in Table 10, we use the information from this step to indicate which feature is observed significantly more among high or low-engaged employees.

Third, we employed Random forest to test if these features can be used in a model to correctly classify employees in the highest 10% or lowest 10% of work engagement. Table 9 presents the models and shows that, compared with study 1, all but one models performed worse. The unigrams, bigrams and linguistic features performed worse (accuracy scores of 52%, 53% and 54%) and barely outperformed a random model. However, the model with psychological features still has an accuracy score of 60%. For comparison, we also conducted Study 2 using the Naive Bayes classifier rather than Random forest. We find that for unigrams the results improve much (from 52% to 64%), whereas for the other features the results are only slightly better (bigrams: 56% instead of 53%; psychological features: 61% instead of 60%; linguistic features: 55% instead of 54%). Appendix 7 presents all results for Naive Bayes.

Fourth, we analyzed the features that best explain high or low work engagement in the different models. For that, we assessed the discriminatory value of the features based on the mean decrease in accuracy of the models when a specific feature is excluded. Table 10 presents the (most) strongly contributing features that were also significantly more present among either high- or low-engaged employees (Appendix 5 presents the overview of all features). Notably, the total amount of contributing features decreased drastically for the unigrams and bigrams, indicating that many features that were used in the first sample were not used in the second sample. In the following discussion section, we compare the features to those of study 1.

Table 9 *Results Random forest for unigrams, bigrams, psychological and linguistic features*

		Model 1: unigrams		**Model 2: bigrams**		**Model 3: psychological features**		**Model 4: linguistic features**	
No. of features		156		24		16		6	
Confusion matrix test set (actual/predicted)									
TP	FP	313	294	383	356	295	202	313	281
FN	TN	134	153	64	91	152	245	134	166
Model statistics									
Accuracy		0.5213		0.5302		0.604		0.5358	
95% CI		0.4879-0.5544		0.4969-0.5633		0.5711-0.6363		0.5025-0.5689	
NIR		0.5		0.5		0.5		0.5	
p-value (Acc > NIR)		0.108		0.038		< 0.001		0.0175	
Kappa		0.0425		0.0604		0.2081		0.0716	
McNemar's Test p-value		< 0.001		< 0.001		0.009		< 0.001	
Sensitivity		0.7002		0.8568		0.6600		0.7002	
Specificity		0.3423		0.2036		0.5481		0.3714	
Pos. Pred. Value		0.5157		0.5183		0.5936		0.5269	
Neg. Pred. Value		0.5331		0.5871		0.6171		0.5533	
Prevalence		0.5000		0.5000		0.5000		0.5000	
Detection Rate		0.3501		0.4284		0.3300		0.3501	
Detection Prevalence		0.6790		0.8266		0.5559		0.6644	
Balanced Accuracy		0.5213		0.5302		0.6040		0.5358	
Hyperparameter values									
Nodesize		1		5		10		5	
Mtry		4		2		5		1	

Table 10 *Contribution of features to the models*

	Model 1: unigrams*	**Model 2: bigrams**	**Model 3: psychological features**	**Model 4: linguistic features**
No. of contributing features				
	39	4	8	2
Most strongly contributing features				
1	pension	**goes well**	**Social processes**	Negations
2	**good**	direct contact	Power (Drives)	**1st person plural**
3	few/little	**we good**	**Positive emotion**	-
4	unsafe	home work	**Reward (Drives)**	-
5	workplace	-	Negative emotions	-
6	**nurtured**	-	Future focus (Time orient.)	-
7	work pressure	-	Work (Personal concerns)	-
8	**resident**	-	**Affiliation (Drives)**	-
9	management	-	-	-
10	**nice**	-	-	-
11	**happily**	-	-	-
12	manager	-	-	-
13	**again**	-	-	-
14	free	-	-	-
15	whereby	-	-	-
16	**our**	-	-	-
17	insufficient	-	-	-
18	leave	-	-	-
19	**sad**	-	-	-
20	unrest	-	-	-

Note. Features that are present significantly more among high-engaged employees are in bold, features present significantly more among low-engaged employees are in normal font. *The features presented in this column are the 20 most strongly contributing features and have a chi^2 -value of 3.9 or higher.

3.6.1. Additional analysis demographics

We conducted additional analyses to investigate how the sample demographics (gender, age, and healthcare branch, as described in Table 3) may affect the results. First, Appendix 6 presents some significant differences in work engagement across gender, age and healthcare branch in both samples. In all cases, effect sizes were small. Second, we explored whether demographics play a role in explaining work engagement. We used the DALEX package (Biecek, 2018) to analyze how gender and age relate to the text features in explaining work engagement. The results (in Appendix 6) show that, next to the text features, gender and age contribute to the models. This suggests that gender and age of healthcare employees contribute to explaining work engagement, and that this may partially confound the effects in our main analysis. A next step in future research would therefore be to add these to the analyses. We discuss further implications in the discussion section.

3.7. General discussion

In this article we aimed to explain work engagement by analyzing self-narratives through text mining. We compared unigrams, bigrams, psychological features and linguistic features. After the explorative approach in study 1, for study 2, we used a confirmatory approach to assess whether the same text features, dependent on both the type of feature and content of the features, can explain work engagement across two samples. From both studies, we deduce three main findings that we want to highlight. First, psychological features can correctly classify healthcare employees into high or low work engagement with 60% accuracy across samples. Second, the features that contribute to the classification partly confirm the literature on the JD-R theory and work engagement. Third, the features also unlock new insights by extending and questioning work engagement theory.

First, we find that the model with psychological features explained work engagement best in both studies, with 60% accuracy. In the first study, unigrams generated the best model (62% accuracy), but the unigrams performed worse in the second study (52% accuracy). This indicates that dictionary approaches, which measure underlying meanings and categories of words, have more success in explaining work engagement than bag-of-word approaches using specific words (Tausczik and Pennebaker, 2010). A likely explanation is that although employees may write about similar topics, they may use different words.

Second, some of the features that contribute to the classification partly confirm the extant literature on antecedents and outcomes of work engagement. Based on study 1, we proposed three hypotheses, supported by the literature, regarding prominent features that explained work engagement. In evaluating our hypotheses, we focus on the model with psychological features, as this is the only model that performed consistently. Drawing conclusions from a model that does not outperform a random model would not be appropriate (we will pay some attention to the model with linguistic features as it still performs slightly better than random). First, we expected that referring to positive emotions contributes to explaining high work engagement, whereas referring to negative emotions contributes to explaining low work engagement (***H1)***. This hypothesis is confirmed in study 2 because, again, positive emotions were more present among high-engaged employees and negative emotions were more present among low-engaged employees. Second, we expected that referring to a crisis contributes to explaining low work engagement, whereas referring to a normal work context contributes to explaining high work engagement (***H2)***. This hypothesis was only based on unigrams and bigrams that referred to the COVID-19 crisis. These models were not able to explain work engagement in the second study. Therefore, hypothesis 2 was not supported. Third, we expected that referring to affiliation contributes to explaining high work engagement (***H3)***. This hypothesis is confirmed too, because again referring to social processes and affiliation explains high work engagement. Besides the hypotheses, three other psychological features contributed across two samples: high-engaged employees referred significantly more to rewards (a dictionary with words like 'benefit', 'bonus' and 'promotion'), and low-engaged employees referred significantly more to power (words like 'manager', 'attack', and 'dependent') and work concerns (words like 'job', 'burden' and 'junior').

Our third finding is that text mining unlocks new insights that extend or question common findings in the literature. First, we are able to uncover that work engagement is related to linguistic patterns. Mainly, across two samples, employees with high work engagement use more first-person plural (e.g., 'we', 'our') than employees with low work engagement. Second, some findings are puzzling and allow to question the literature. For example, in the first study, there are multiple unigrams that refer to management and across two samples, there is a psychological feature that refers to power. What is striking is that these features all contribute to explaining low work engagement, suggesting employees who are low-engaged tend to mention their managers more. We also observe features that refer to certain subgroups of employees. For example, in both samples, the unigram 'retirement' contributes to explaining low work engagement. Exploratory analyses of self-narratives that include this unigram suggest that these are employees who are close to retirement. And in sample 1, the unigram 'caregiver' contributes to explaining high work engagement. Exploratory analyses suggest that these employees are voluntary caregivers besides their regular work. We should be careful to interpret these exploratory findings, and we provide potential explanations below.

3.7.1. Scientific and practical implications

Our findings have multiple implications. Regarding implications for theory, our study further the understanding of work engagement as a theoretical concept. First, text mining enables validation of findings in the extant literature and complements these findings with rich context due to a large-scale analysis of self-narratives. As we explained in our theory section, antecedents of work engagement are often studied within the JD-R theory, a theory within organizational psychology that explains how job characteristics affect employees through a dual process. Job resources foster a motivational process leading to positive outcomes like work engagement, whereas hindrance job demands cause a health impairment process and diminish the positive effects of job resources on work engagement (Van Veldhoven et al., 2020; Bakker and Demerouti, 2017; Schaufeli et al., 2009). Our results confirm this duality because the features describe resources and demands. The features that high-engaged employees refer more often to, like affiliation and rewards, are often job resources (Bakker, 2022; Wang, Zhu et al., 2018; Bakker et al., 2014). For example, experiencing social support positively affects work engagement (Freeney and Fellenz, 2013). Even more so, work engagement can be a truly contagious process transferring between employees (Bakker et al., 2016), and even from employees to partners (Bakker et al., 2005) and home life (Culbertson et al., 2012). Likewise, positive emotions, another feature more present among high-engaged employees, can be considered personal resources (Ouweneel et al., 2012). Contrarily, the features that low-engaged employees refer to, like power and work concerns, tend to be job demands. Power refers to words describing hierarchy or dependency, with words like 'manager', 'attack', and 'dependent'. This resembles studies that have shown that abusive supervision or bullying is negatively related to work engagement (Wang, Hsieh et al., 2020; Einarsen et al., 2018), and may also point to the absence of autonomy, an important resource and antecedent of work engagement (Christian et al., 2011). Finally, negative emotions, another feature more present among low-engaged employees may suggest that emotional instability be regarded as a personal demand that affects work engagement (Lorente Prieto et al., 2008).

Second, whereas some findings confirm the duality of JD-R theory, we also found remarkable linguistic patterns that extend it, and relatively unexplored antecedents of work engagement that question it. These findings can increase our understanding of work engagement and how it is theorized and measured. First, the finding on linguistic differences between high and low-engaged employees uncovers a new research area that may focus on work engagement markers within speech or writing. Until now, studies have mostly focused on linguistics in more clinical concepts, like schizophrenia (Franklin and Thompson, 2005). Studying linguistic patterns may increase our understanding of work engagement, especially in the context of diary studies, as these studies allow employees to provide unstructured data on a regular basis (Zampetakis, 2023; Ouweneel et al., 2012). Specifically, the finding that high-engaged employees use more first-person plural is a tangible indication of the social and contagious nature of work engagement (Bakker, 2022). Second, in the self-narratives low-engaged employees more often referred to their managers. This finding is puzzling. The literature shows that managers can have important, positive influences on employee well-being and often finds positive effects of 'good' leadership styles or behaviors on work engagement (Tummers and Bakker, 2021; Decuypere and Schaufeli, 2020). In contrast, managers can also have negative influence, when they bully or execute abusive supervision (e.g., Barnes et al., 2015). Our results suggest that employees are more likely to mention managers if they are a negative influence. A potential explanation is that positive behaviors are more seen as a self-evident part of a managers' role (Toegel et al., 2013). COVID-19 has been a tremendous leadership challenge (Graham and Woodhead, 2021), and especially employees who experienced failing leadership may have wanted to mention this in their self-narratives. In any case, the results suggest a vital role for managers in fostering employee work engagement (Freeney and Fellenz, 2013). Third, some findings beg for further research. For example, a recent study suggests that 'mental retirement' among older employees is non-existent (De Wind et al., 2017). At the same time, in our study employees who were low in engagement more often referred to retirement. One explanation is that working during COVID-19 has been especially burdensome for older employees (Van Roekel et al., 2021). This emphasizes the need for interventions that support older employees in the workplace (Söderbacka et al., 2020). In contrast, the finding that high-engaged employees refer more to being a voluntary caregiver besides their work points to another avenue in which work engagement may affect home life and cause citizenship behavior (Culbertson et al., 2012; Xanthopoulou et al., 2008).

Our main methodological contribution is that, to our knowledge, this is the first study that succeeds in explaining work engagement by text mining self-narratives. The best-scoring model in the first sample uses unigrams (62% accuracy) and the best-scoring model across samples uses psychological features (60% accuracy in the second study). We argue that our study indicates that, for work engagement, classification by text mining cannot easily replace structured forms of data analysis as it is not precise enough yet. Nevertheless, there are multiple avenues in which text mining could support and complement more traditional data analysis. First, it could validate the relative importance of antecedents and outcomes of work engagement. For example, if a relationship between work engagement and another concept, for example, empowering leadership, is analyzed, additional text mining of open questions could indicate differences in the way employees in high or low categories of work engagement discuss their managers (e.g., Tuckey et al., 2012). Text mining could also perform a supporting role by being used in the validation of scales, for example, by analyzing what words employees use to describe being engaged at work. Scales that use this as input for wording may be more ecologically valid (Kobayashi et al., 2021). Text mining could also complement structured forms of data analysis by using it as an exploration of what topics and concepts are associated with work engagement but may have received little attention in the literature. Besides, in situations where lengthy surveys are not preferred, text mining enables efficient analysis of an open question (Kobayashi et al., 2021; Jurafsky and Martin, 2017).

Finally, and this is both a methodological and practical implication, text mining may present a new avenue for *in vivo* assessment of work engagement. Now that this study has found that survey-based self-narratives explain work engagement to some extent, future research could use existing data to attempt to do the same. Albeit for scientific or managerial purposes, existing texts (like shared diaries or intranet posts) or other forms of unstructured data within organizations may very well allow for the screening and identification of employees whose work engagement is challenged (e.g., He et al., 2012; Day et al., 2007). In addition, studies could employ text mining techniques to present employees with a self-assessment of work engagement. Employees could, after providing a self-narrative, perhaps receive a comparative score and/or a personalized suggestion, like talking to a confidant. By making assessment easier, text mining could perhaps be a preventive HR tool, if employee privacy is maintained and the interest of employees is put first. Besides, the exploration of the features that contribute to explaining work engagement may help employees, (HR) managers, and (healthcare) organizations to more quickly recognize and act upon challenges to work engagement. The features that turned out to be important may indicate resources where organizations should invest in, like guaranteeing adequate social support systems and stimulating social contact between employees. Likewise, organizations should pay attention to employees' emotional state. Gauging healthcare employee work engagement has become increasingly relevant since the COVID-19 crisis, which has been challenging especially among healthcare employees dealing with COVID-19 patients (Van Roekel et al., 2021).

3.7.2. Limitations

There are limitations to this study. First, the data we used present limitations. We compare self-narratives to work engagement scores within the same survey, which may lead to common source bias. Using two survey waves has increased the strength of our design. Still, future research could go beyond survey-based analysis by employing human coders (e.g., psychologists) to assess self-narratives. Likewise, our text mining data were survey-based and created specifically for this study. This somewhat limits the external generalizability of our findings when discussing opportunities for text mining of existing, unstructured data. It is common, however, to begin with manufactured data and then expand to pre-existing data after (e.g., He, 2013). Therefore, future research could use such pre-existing data like email, intranet, or social media messages to address this limitation (e.g., Pang et al., 2020). Finally, the survey did not use attention checks, which may be regarded as a limitation. Nevertheless, there is considerable discussion in the literature about their effectiveness and necessity. Recent findings suggest attention checks do not harm scale validity but removing those who fail attention checks often does not alter substantive analyses either (e.g., Gummer et al., 2021; Kung et al., 2018).

Second, there are limitations related to sample characteristics. The dataset is unbalanced because there are more employees who score high versus low on work engagement (high work engagement: M = 4.87 for sample 1 and M = 4.84 for sample 2; low work engagement: M = 2.60 for sample 1 and M = 2.46 for sample 2). This limitation indicates that our analysis strictly explains the differences between very high work engagement and work engagement lower than the midpoint (i.e., 3) of the scale. One explanation is that healthcare employees are generally high in work engagement, so the limited generalizability of our results may be more pronounced in sectors with lower work engagement, like manufacturing (Hakanen et al., 2019). Nevertheless, future research may explicitly include employees with low work engagement by, for example, targeting employees who intend to quit their jobs (e.g., in exit interviews).

Another limitation regarding sample characteristics is that our main analysis focused on text-based features and therefore ignored the role of demographics such as gender and age. However, the literature shows work engagement can vary depending on gender and age, (although only to a limited extent; see e.g., Schaufeli et al., 2006). Yet we also argued that gender and age may affect the results because of the particularities of our sample and, to some extent, the population of healthcare employees. Hence, we can expect gender and age to meaningfully relate to work engagement in our particular samples. Controlling for gender and age in additional analysis confirms that these variables do play a role. Although these findings should be taken into account, our goal was not to develop the best model to explain work engagement but the best fitting model with text features from a representative sample of healthcare employees. Having a fairly representative sample for healthcare is a strength of our study's generalizability within healthcare and comparable sectors, but does limit generalization when it comes to sectors with different characteristics. With this restriction in mind, our results contribute to understanding what text-based features contribute to explaining work engagement. Future research may extend our findings by paying more attention to the role of demographics in text mining research and by repeating our methods in different contexts.

A final limitation regarding sample characteristics concerns differences in respondent characteristics between the two samples. In comparison with the first sample, the second sample contains fewer respondents who also wrote shorter texts and had a lower mean work engagement. A potential explanation for the difference in participation rates is respondent fatigue: respondents may have been more motivated to provide a self-narrative when the request for such a narrative was newly introduced compared with when it was repeated. However, the drop in work engagement may also point to another explanation: as the COVID-19 crisis continued, healthcare employees were exposed to persistent job stress, which may have caused the decrease in general levels of work engagement between the 2020 sample (#1) and the 2021 sample (#2) (Kniffin et al., 2021; Van Roekel et al., 2021; Wang et al., 2021). Sample heterogeneity may have affected the translatability of text features across samples somewhat and may have decreased the reliability of the models. However, we did find that the samples were comparable when it comes to gender, age and healthcare branch. Still, future research could attempt to collect samples with identical respondent characteristics to counter sample heterogeneity.

Third, limitations apply regarding the methods used. This study aimed to explore the possibility of text mining for work engagement classification using Random forest and Naive Bayes. Our results showed that with Random forest we were able to classify, but in Study 2 Naive Bayes performed better than Random forest. This may inform future use of classifiers for text-based features. Yet there are more possibilities for future research and further optimization of the methodology. Other approaches, including statistical methods such as LASSO feature elimination and OLS regression, and machine learning methods such as support vector machines (SVMs), could prove better suited to the data at hand. This should be decided on a case-by-case basis depending on the data and project goals. For our project, we primarily used Random forest as it presents feature importance information, which allows us to understand what features contribute most to the models. Besides, compared with regression models, it is able to handle the high dimensionality of text data better. Additionally, compared with regression, Random forest is more robust to outliers. Finally, Random forest is also suited for non-linear relationships and categorical variables. SVMs share some of the advantages of Random forest, but are heavier and harder to interpret. In sum, following up on our study, researchers could employ a variety of methods to provide new insights into the uses of text mining.

Fourth, our results are promising but the models are not nearly 100% reliable. One of the reasons may be that the self-narratives were relatively short (compared with e.g., He et al., 2012). Longer stories may lead to better explanations. Besides, there is the issue of the ‘middle 80%’: we cannot readily make statements about all respondents in between the highest or lowest 10%. Our approach is a most-likely-case scenario, if we do not find differences between these two groups, there most likely will not be any differences found for the 80%. If we do find differences between these two groups, these differences will most likely be more pronounced than the differences for the 80%. Our recommendation for future research is to look beyond binomial categorization. Different feature selection methods, data representations or neural network approaches to text classification could improve model performance further. Likewise, taking inspiration from our approach to studying work engagement, scholars could expand our study and include different features to test their respective contributions to explaining work engagement. By doing so, scholars can continue to confirm, extend and/or question the literature. For example, extension could take place by studying unexplored features. We also see opportunities to further question the literature if scholars find features that contribute more to the reliability of the models than the features that we studied, or if features suggest contrary relationships between work engagement and other variables compared with our results or established theories on work engagement. Besides, we explored a bag-of-words and dictionary approach to text classification for work engagement. This means that syntactic and contextual information is not taken into account. Modern approaches that focus on further understanding relationships in the text may help future research do enrich the analysis. Word embeddings such as Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) allow to keep syntactic information intact instead of considering each word as a standalone feature. Recently, approaches such as BERT (Devlin et al., 2018) and OpenAI GPT (Radford et al., 2018) take this even further by incorporating contextual knowledge in the model using pretraining.

Fifth, we only addressed the concept of work engagement. It remains a question of how other measures, like positive and negative affectivity, compare with our text mining approach. Future research could explore the comparative explanatory accuracy of such measures in the work context.

Finally, the most important limitation regarding the results is inherent to text mining: ‘text mining procedures in and of themselves cannot support causal inference (i.e., internal validity) unless the study design is such that, next to association, temporal precedence and isolation are also established’ (Kobayashi et al., 2021, p. 148). We analyzed associations, not causal relationships.

3.7.3. Conclusion

In this paper, we aimed to introduce text mining as a methodological approach to study employee work engagement, and, more generally, text mining as a method in organizational research. Our study attempted to analyze work engagement, and the features that contributed to the models help explain what it means to be (or not to be) engaged in work. Text mining truly allows to assess the multidimensionality of a phenomenon (Balducci and Marinova, 2018), and so, like qualitative research, it offers a richer description of reality, but, like quantitative research, it is able to handle large amounts of data. In sum, text mining is an interesting and innovative approach that may be used to validate but also complement findings from studies with more structured approaches to studying work engagement.

Chapter 4

Empowering leadership during a crisis

This chapter is based on the following manuscript:
Van Roekel, H., Sieweke, J., Schott, C., Bakker, A. B., & Tummers, L. G. The effect of empowering leadership on employee well-being during a public health crisis: a natural experiment. *Under review.*

Abstract

While studies show empowering leadership improves employee well-being, empirical research in public context is limited. We investigated whether the positive effect of empowering leadership on employee well-being would differ during a public health crisis. We designed a natural experiment, using a longitudinal survey among Dutch healthcare employees and administrative data on the geographical variance in COVID-19 hospitalization rates. The findings show that empowering leadership is less effective in a crisis and may even harm employee well-being. Our study questions the proposition that empowering leadership is always beneficial for employee well-being and contributes to understanding its dark side in public context.

4.1. Introduction

Managers in both the public and private sector increasingly consider transferring influence to employees through empowering leadership (e.g., Kang et al., 2022; Amundsen and Martinsen, 2014; Parker, Wall and Corderly, 2001). Whilst scholars were initially particularly interested in the effects of empowering leadership on employee performance, nowadays they also focus on its relationship with employee well-being (Kim, Beehr and Prewett, 2018). This mirrors a more balanced approach towards Human Resource Management, arguing that both well-being and performance are important outcomes in organizations (e.g., Amundsen and Martinsen, 2015; Park et al., 2017; Van de Voorde, Paauwe and Van Veldhoven, 2012). Two recent meta-analyses (Kim, Beehr and Prewett, 2018; Lee, Willis and Tian, 2018) show empowering leadership is generally associated with higher well-being, for example, higher work engagement (Schaufeli et al., 2002).

Whereas most studies find evidence for the positive side of empowering leadership, some studies provide evidence for a dark side that decreases well-being (e.g., Dennerlein and Kirkman, 2022; Cheong et al., 2016). To explain these conflicting findings, one particular question may be relevant: when does empowering leadership (not) work? In other words, what are contextual factors that affect the effect of empowering leadership on well-being? We know little about in what contexts empowering leadership improves—or maybe even harms—well-being (Kim, Beehr and Prewett, 2018; Sims Jr., Faraj and Yun, 2009). This is problematic as leadership does not exist in a vacuum: paying attention to heterogeneity in effects due to contextual factors is vital to understanding the potential of empowering leadership (Porter and McLaughlin, 2006). In this study, we contribute to this debate by addressing an important contextual factor: a crisis.

In a crisis, organizations experience high work intensity through urgent and highly uncertain threats to their 'core values or vital systems' ('t Hart and Tummers, 2019, pp.120-121; Yun, Faraj and Sims Jr., 2005; Tuckey, Bakker and Dollard, 2012). In the present study, we focus on a public health crisis caused by the outbreak of the COVID-19 crisis and its effects on the work and well-being of public employees (Schuster et al., 2020), especially those dealing with COVID-19 patients in the healthcare sector (Van Roekel et al., 2021). These healthcare employees have been confronted with increasing job demands, like patient loads, and decreasing job resources, like supply shortages (Kniffin et al., 2021). This impoverished job design has negatively affected employee well-being in the form of, among others, increased stress, anxiety, depression, and insomnia (Spoorthy, Pratapa and Mahant, 2020).

Unfortunately, empirical research on employee empowerment in crises is limited, especially in the public sector ('t Hart and Tummers, 2019). To formulate our hypotheses on the effect of empowering leadership on well-being in a crisis, we build on Job Demands-Resources (JD-R) theory (Borst and Knies, 2021; Bakker and Demerouti, 2017; Kim, Beehr and Prewet, 2018). The basic and frequently studied argument is that empowering leadership provides employees with resources like autonomy and psychological capital, which improve well-being (Park et al., 2017). In addition, the JD-R literature suggests that job resources are especially effective in improving well-being when demands are high (Bakker et al., 2007). Together, this suggests empowering leadership may improve well-being especially in a crisis (Bakker et al., 2014).

With our research, we aim to answer the question: does the effect of empowering leadership on employee well-being differ during crisis? To do so, this study exploits the geographical differences in the intensity of the COVID-19 crisis in the Netherlands, which provides opportunities for estimating the effect of varying degrees of treatment intensity by applying a difference-in-differences (DID) design.

Our study aims to make several contributions to the literature. The empowering leadership literature provides promising evidence on how leaders may stimulate motivation and work success of their employees through empowerment (Cheong et al., 2019). Yet, empowering leadership studies tend to neglect the role of context, and we test whether empowering leadership is a positive force in the context of a crisis (Sims Jr., Faraj and Yun, 2009). Additionally, in the public leadership literature on crises we find an increasing interest in empowerment as an avenue for leaders to deal with crises (Boin and 't Hart, 2003). However, its effects are little empirically researched, especially in public management ('t Hart and Tummers, 2019). Therefore, our study contributes to the public leadership literature by testing to what extent empowerment is a solution during a crisis, considering employees' well-being as a factor. In addition, for practitioners, this study offers a nuanced understanding of the potential of empowering leadership that may help practitioners execute situational leadership.

4.2. Theoretical background

4.2.1. The effect of empowering leadership on employee well-being

The nature of work has, in the last decades, become more complex and those executing the work have become increasingly skilled (Parker, Wall and Corderly, 2001). In response, management scholars engaged in studying employee empowerment as a way to increase employee productivity (Bartunek and Spreitzer, 2006). It was proposed empowerment would enhance motivation through delegating responsibility and authority (Conger and Kanungo, 1988, Thomas and Velthouse, 1990). To foster empowerment among employees, managers can portray empowering leadership, a set of behaviors that focus on the transfer of power and influence (Kang et al., 2022; Amundsen and Martinsen, 2014). According to Ahearne, Mathieu, and Rapp (2005), managers portraying empowering leadership a) enhance the meaningfulness of work, b) foster participation in decision making, c) express confidence in high performance, and d) provide autonomy from bureaucratic constraints.

Researchers have paid increasing attention to empowering leadership and its outcomes in both public and private sector contexts (Chen et al., 2023). Most importantly, they focused on the relationship between empowering leadership and employee performance (e.g., Park et al., 2017; Kim and Holzer, 2016). For instance, a meta-analysis by Lee, Willis and Tian (2018) finds that empowering leadership is positively related to individual task performance, organizational citizenship behavior and creativity, as well as team-level performance.

More recently, scholars emphasize employee well-being as a vital outcome of leadership, hereby mirroring Human Resource Management's balanced approach emphasizing the importance of both well-being and performance (e.g., Amundsen and Martinsen, 2015; Van de Voorde, Paauwe and Van Veldhoven, 2012). Employee well-being refers to the 'overall quality of an employee's experience and functioning at work' (Warr, 1987; in Grant et al., 2007, p.52). So far, the literature indicates a positive association with employee well-being. A meta-analysis by Kim, Beehr and Prewett (2018) found positive associations with employee motivation and psychological resources (e.g., psychological empowerment) and with employee attitudes (e.g., work engagement).

4.2.2. Deepening the empowering leadership-well-being relationship: contextual factors

To explain how empowering leadership affects well-being, most studies use JD-R theory (Kim, Beehr and Prewet, 2018). A core argument of JD-R theory is that in their work employees are confronted with job characteristics or events that cost effort and energy (job demands), and job characteristics or events that are enriching and motivating (job resources). These demands and resources instigate two separate processes: one that decreases health and leads to exhaustion (i.e., a health impairment process), and one that improves motivation and leads to engagement (i.e., a motivational process) (Borst and Knies, 2021; Bakker and Demerouti, 2017; Bakker et al., 2014). Several studies show empowering leadership contributes to well-being by generating resources (Tummers and Bakker, 2021). For example, Tuckey, Bakker and Dollard (2012) find that increases in cognitive demands and resources partly mediated the relationship between empowering leadership and work engagement, and that empowering leadership optimizes working conditions for engagement. The positive association of empowering leadership with work engagement can, similarly, be explained through the resource of autonomy: empowering leaders increase autonomy, which is a job resource that subsequently impacts vitality and work engagement (Albrecht and Andreetta, 2011). Finally, empowering leadership contributes to job crafting (Audenaert et al., 2020).

Yet, some studies also highlight a dark side of that empowering leadership (Dennerlein and Kirkman, 2022; Sharma and Kirkman, 2015). For instance, Dennerlein and Kirkman (2022) find empowering leadership can increase unethical behavior through increasing moral disengagement. Focusing on well-being, Cheong et al. (2016) find empowering leadership does not only improve employees' self-efficacy through an enabling process; it also increases job induced tension through a burdening process. The burdening process shows that greater autonomy may come at the cost of increasing strain (Langred and Moye, 2004), and that the added responsibilities could lead to role stress (Kahn et al., 1964). In sum, whilst most studies focus on and find evidence for the positive side of empowering leadership, some studies provide evidence for negative effects of empowering leadership.

A crucial question that results from these conflicting findings is when empowering leadership works (i.e., improves employee well-being) and when not (i.e., does not improve or even decreases well-being). That is, what are contextual factors that moderate the effect of empowering leadership on well-being? This question is very pressing because the full potential of a leadership style cannot be understood without paying attention to heterogeneity in effects due to contextual factors (Porter and McLaughlin, 2006). In the leadership literature, scholars commonly agree that a specific type of leadership might be best for a specific situation (e.g., Stoker, Garretsen and Soudis, 2019; Sims Jr., Faraj and Yun, 2009). However, in the empowering leadership literature, the role of context is understudied (Sims Jr., Faraj and Yun, 2009). For example, there is little, contrasting research on the boundary conditions of work intensity and its effect on the effectiveness of empowering leadership (Yun, Faraj and Sims Jr., 2005; Tuckey, Bakker and Dollard, 2012). In this study, we aim to address this gap by focusing on the effects of empowering leadership on employee well-being in times of crisis, i.e., under high work intensity.

4.2.3. Empowering leadership and employee well-being during a crisis

The public leadership literature indicates that crises require extraordinary leadership skills for organizations or societies to stand a fair chance (Boin et al., 2016). A crisis is ‘a serious threat to the basic structures or the fundamental values and norms of a system, which under time pressure and highly uncertain circumstances necessitates making vital decisions’ (Rosenthal et al., 1989, p.10). An important lesson from research on crises and high-reliability organizations is that empowerment of employees and lateral coordination is a crucial strategy: expertise rather than hierarchical position should be leading (e.g., Boin and ‘t Hart, 2003). Scholars studying employee resilience (the individual quality to deal with adversity) found this could be supported by empowerment- and autonomy-based job designs (McDonald et al., 2016). However, empirical research on the consequences of empowerment in crises is limited, particularly in the public sector (‘t Hart and Tummers, 2019). Therefore, to hypothesize about the relationship between empowering leadership and well-being in a crisis, we consider, first, the specific crisis we study, and second how this crisis would affect the relationship of empowering leadership and well-being.

In our study, we focus on the COVID-19 pandemic. This crisis affected employees in healthcare that suddenly had to work with COVID-19-infected patients (Van Roekel et al., 2021; Schuster et al., 2020). Many healthcare employees found themselves in the front line of the battle against the virus. Consequentially, a stream of literature on the effects of COVID-19 on healthcare work and employee well-being is rapidly evolving (e.g., Kniffin et al., 2021). It shows the COVID-19 pandemic confronted many healthcare employees with increasing demands and decreasing resources: resources decreased, as there were critical supply shortages, like ventilators and personal protective equipment (Ranney, Griffeth and Jha, 2020); demands increased as patient loads increased whilst safety measures and precautions, like wearing protective equipment, increased the complexity of the work. At the same time, the COVID-19 pandemic also marked an increase in the relevance of, and appreciation for, the work of healthcare employees (Shand et al., 2022; Mohindra et al., 2020; Bagcchi, 2020), as illustrated by people applauding healthcare workers.

This raises the question how the ongoing COVID-19 crisis may influence the effect of empowering leadership on well-being. We apply JD-R theory to answer this question. Multiple studies show job resources are especially effective in improving work engagement when job demands are high (Bakker et al., 2007). This can be explained by the dualistic nature of resources and demands: ‘high levels of job demands provide the conditions to motivate workers for action, and high levels of job resources provide the means to carry through with the action plan’ (Tuckey, Bakker and Dollard, 2012, p.17). Consequently, we expect the COVID-19 crisis to positively moderate the positive impact of empowering leadership on well-being, as the crisis presents a combination of resources, like autonomy and psychological capital, with high demands (Tummers et al., 2020; Park et al., 2017; Albrecht and Andreetta, 2011). Herein, we follow the findings from the majority of the studies on empowering leadership (see meta-analyses; Lee, Willis and Tian, 2018; Kim, Beehr and Prewett, 2018).

Finally, we expect effects differ somewhat for the energetic dimension (experiencing vigor) of employee well-being compared to the cognitive and motivational dimensions (being dedicated to, and absorbed in, work). Recent reviews show that the increase in COVID-19-related job demands, like increased patient loads, have led to more stress, anxiety, and insomnia among healthcare workers (Spoorthy, Pratapa and Mahant, 2020; Shreffler, Petrey and Huecker, 2020). At the same time, preliminary evidence suggests increased job resources, like appreciation and support, boosted healthcare workers' morale and job pride (Mohindra et al., 2020). These job resources may help employees counter the increase in demands. However, studies show that the energetic dimension of well-being is more prone to a health impairment process caused by increased demands and stressors than the motivational and cognitive dimensions (Bakker and Demerouti, 2017). Consequently, employees' energetic dimension may be more challenged and empowering leadership may be as more necessary for one's energy level compared to one's cognitive and motivational levels. In sum, the positive effect of empowering leadership during crisis increases for all dimensions of employee well-being, but more so for the energetic dimension.

Our key hypothesis summarizes this argument:

H1: *The positive effect of empowering leadership on healthcare employee well-being will be positively moderated by a crisis.*

Specifically,

H1a: *The positive effect of empowering leadership on the motivational and cognitive dimensions of healthcare employee well-being increases during a crisis.*

H1b: *The positive effect of empowering leadership on the energetic dimension of healthcare employee well-being increases during a crisis, more so than the effect on the motivational and cognitive dimensions.*

4.3. Methods

4.3.1. Research Context, Data and Sample

This study focuses on the COVID-19 pandemic. The outbreak started in December 2019 and within about 3 months, the virus spread around the world. On March 11, 2020, the World Health Organization announced that the COVID-19 outbreak could be characterized as a pandemic.

For this study, we collected data in a longitudinal survey study on leadership and well-being of Dutch healthcare employees. Data collection for the surveys was approved by the Faculty Ethical Review Committee of Utrecht University. The data could not be made available publicly due to privacy regulations but are available upon reasonable request and after signing a processing agreement by emailing the corresponding author. We preregistered this study at the Open Science Framework (see List of preregistrations and Appendix 2).

The study was executed among members of IZZ, a healthcare employee collective. For this study, we used data collected in two waves: March-May 2019 and May-June 2020. The data collections took place respectively before and after the so-called ‘first wave’ of the COVID-19 pandemic in the Netherlands. From March until May 31, 2020, the Netherlands reported ca. 46,000 infections with ca. 11,700 hospital admissions and ca. 5,900 deaths out of a population of around 17 million inhabitants (RIVM, 2020).

A total of 1,379 respondents who had participated in both the 2019 (i.e., before the crisis) and 2020 wave (i.e., during the crisis) were linked through a Self-Generated Identification Code (SGIC; Schnell et al., 2010). A Python toolkit was used to link the datasets (De Bruin, 2019). Given our focus on how a crisis moderates the relationship between empowering leadership and well-being of healthcare employees and the large variety of healthcare-related jobs performed by our respondents (e.g., policy officer or IT advisor in healthcare), we focused only on respondents who confirmed they had been in contact with COVID-19 patients, because these workers directly experienced the consequences of the crisis. Our final sample consisted of 468 respondents for whom we had observations at two points of time (2019 and 2020). Thus, our final sample size was 936 observations.

Of the 468 respondents, 86.7% were female, which is similar to the general healthcare population (84.3%, see CBS, 2020a) and 41.7% were 55 years or older, which is considerably higher than their share among healthcare workers (24.2%, see CBS, 2020a). Our respondents represented the major healthcare industries: hospitals (59%, 21.9% in population), nursing homes and homecare (25%, 32.1% in population), mental healthcare (5.1%, 7% in population), disability care (6.4%, 13.4% in population) and other (4.5%) (CBS, 2020a). The overrepresentation of hospitals was due to the selection criterion of direct contact with COVID-19 patients.

4.3.2. Measures

A Dutch translation of the sample items used in this study can be found in Appendix 1. For each latent construct, we report reliabilities both for data collected in 2019 and 2020 below.

Employee well-being

We measure three dimensions of employee well-being: motivational, cognitive and energetic, with a total of 5 measures (Bakker and Demerouti, 2017). The motivational and cognitive dimensions are measured with the work engagement dimensions dedication and absorption, respectively (Schaufeli et al., 2006). Both dimensions were measured with three items on a 5-point Likert scale ranging from 'Never' (1) to 'Always (daily)' (5). An example item of dedication is: 'I am proud on the work that I do'. The reliability of this scale was good (2019: $\alpha = .82$; 2020: $\alpha = .82$). An example item of absorption is: 'I feel happy when I am working intensely'. The reliability of the scale was good (2019: $\alpha = .82$; 2020: $\alpha = .79$).

The energetic dimension is measured with one dimension of work engagement: vigor; and two additional burnout measures adapted from the MBI-GS: physical exhaustion and mental exhaustion (Schaufeli et al., 1996). We added the two burnout measures as an extra check on our measure of vigor, as scholars have shown there are cases in which especially these variables may diverge (Mäkikangas et al., 2017). Again, all were measured with three items on a 5-point Likert scale ranging from 'Never' (1) to 'Always (daily)' (5). An example item of vigor is: 'At my work, I feel bursting with energy'. The reliability of this scale was good (2019: $\alpha = .84$; 2020: $\alpha = .84$). An example item of physical exhaustion is: 'I feel physically exhausted because of my work'. The reliability of this scale was good (2019: $\alpha = .87$; 2020: $\alpha = .88$). An example item of mental exhaustion is: 'I feel mentally exhausted because of my work'. The reliability of this scale was good (2019: $\alpha = .88$; 2020: $\alpha = .90$).

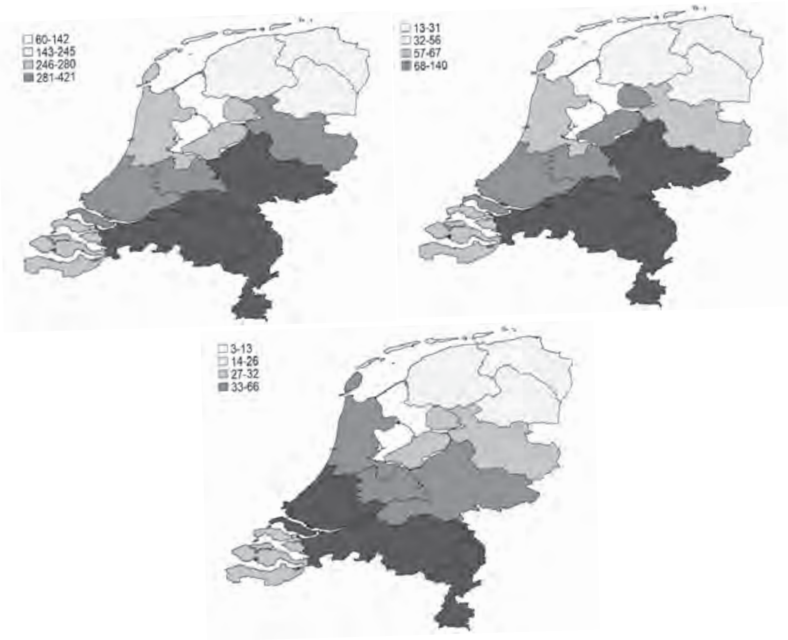
Crisis

We used two variables to operationalize the crisis context. First, we included the dummy variable 'crisis' that indicates whether the respondent was surveyed before (coded '0') or during the outbreak of the COVID-19 pandemic (coded '1'). As we explain in more detail in our statistical analysis section, this variable is crucial for our DID design, because it indicates the treatment period.

Second, we use a continuous measure, which we refer to as 'crisis intensity', that measures the extent to which the COVID-19 pandemic affected the work of healthcare employees in the Netherlands. We operationalized 'crisis intensity' as the COVID-19 hospitalization rate per 100,000 inhabitants within each province of the Netherlands. We summarized all COVID-19 related hospital admissions for the time period from March 13, 2020, until May 31, 2020, as most respondents submitted their survey in May and early June 2020. Therefore, our measure approximates the extent to which COVID-19 affected the jobs of the healthcare employees. We collected the information on the COVID-19 hospitalization rates for each of the twelve provinces from the National Institute for Public Health and the Environment (RIVM). The variable exploits the fact that healthcare employees were heterogeneously affected by the COVID-19 pandemic depending on the Dutch province where they worked. That is, we expect that job demands were much higher for healthcare employees working in provinces with higher COVID-19 hospitalization numbers compared to provinces with lower hospitalization numbers. We expect that hospitalization is a better proxy for job demands than, for instance, COVID-19 infections, because the rate to which infections translate into actual patient treatment at the hospital may differ between provinces due to, for instance, population demography (e.g., age structure).

The RIVM published a daily update of the COVID-19 infections, hospitalization and deaths counts (RIVM, 2020). The data showed a consistent picture: the provinces in the south-east of the Netherlands (i.e., Brabant, Limburg and Gelderland) were affected much stronger by COVID-19 regarding the total number of cases (Figure 1a), hospitalization ratio (Figure 1b) and deaths (Figure 1c) than the northern provinces (i.e., Drenthe, Friesland and Groningen) (RIVM, 2020). For instance, COVID-19 related hospital admissions per 100,000 inhabitants between March and May 2020, were about ten times higher in Limburg (139 admissions per 100,000 inhabitants) than in Groningen (13 admissions per 100,000 inhabitants). Therefore, we can assume that healthcare workers were heterogeneously affected by the crisis depending on the province they worked in.

Figures 1a-c *COVID-19 effects in the provinces of the Netherlands from March until May 2020. Figure 1a (upper left): COVID-19 cases (per 100,000 inhabitants). Figure 1b (upper right): COVID-19 related hospitalization (per 100,000 inhabitants). Figure 1c (lower): COVID-19 related deaths (per 100,000 inhabitants)*



Although the COVID-19 hospitalizations represent a good proxy of the job demands of healthcare workers within a province, the measure may be affected by strategic decisions. First, hospitals may, if needed, have relocated COVID-19 patients to hospitals in provinces with lower COVID-19 hospitalizations to decrease job demands. Yet, we have reasons to believe that this behavior has little influence on the COVID-19 hospitalization numbers. On the one hand, these relocations were included in the hospitalization measures. On the other hand, relocations were primarily done within one of the eleven care regions of the Netherlands, and these care regions are distributed either along province lines or similar geographical distribution (Nederlandse Zorgautoriteit, 2017). Only if none of the hospitals within a care region could help, a patient would be relocated to another care region (LCPS, 2021). This was primarily the case for ICU patients. For example, data from the National Coordination Centre for Patient Distribution (LCPS) show that at the peak of the 'first wave' of the COVID-19 pandemic on April 9, 2020, 1,417 COVID-19 patients were staying on an ICU, but that day only 22 ICU relocations (1.6%) were organized between regions. Also, during the time of our study relocations between the regions decreased quickly: from April 12 onwards, ICU relocations were consistently below 10 and eventually (from May 21 on) became zero (LCPS, 2020). Therefore, we argue that relocations of patients do not considerably affect our COVID-19 hospitalization measure.

Second, the crisis intensity variable could be somewhat affected by the fact that provinces with high crisis intensity increased their staff size, especially on ICUs that had to deal with the most severe cases. The Dutch scaling up plan for COVID-19 provides information on how ICU staff size is increased in the several stages of crisis intensity. In the first stage, current ICU employees work more hours and former ICU employees are invited. The second stage includes the former with addition of other specialized nurses, and the third stage includes the former with addition of other acute nurses and voluntary revoke of leave (LNAZ, 2020). Importantly, any increase in staff size is preceded and accompanied by an increased demand on the current staff. This shows that our crisis intensity variable is a good proxy of job demands even if staff size is increased.

Third, hospitals may relocate general healthcare staff to deal with increased job demands. Indeed, relocation of staff within healthcare organizations was common practice: 35% of all respondents in the 2020 wave of our survey that had direct contact with COVID-19 patients indicated having worked in other wards (n = 2,629). However, the Dutch scaling up plan for COVID-19 notably does not include relocating staff between organizations, let alone provinces (LNAZ, 2020). This finding further shows that our crisis intensity variable is a good proxy of job demands during the COVID-19 pandemic.

Finally, we should consider the extent to which COVID-19 has caused organizational changes that may interfere with our measurement of empowering leadership (Kniffin et al., 2021). In Appendix 3, we address a potential endogeneity threat by assessing whether crisis intensity affects empowering leadership. Besides, an increase in working from home (WFH), especially for managers, may have affected employees' perceptions of empowering leadership (Fischer et al., 2022). However, data from the Dutch Central Bureau for Statistics (CBS) show that for employees with managing or healthcare jobs, WFH did not increase much (for managers: 2019: 38%, 2nd quarter of 2020: 38.8%; for healthcare jobs: 2019: 13.8%, 2nd quarter of 2020: 18%). What is more, we measured empowering leadership for direct supervisors, and CBS data show managers in lower echelons of the organization worked from home less compared to managers with more responsibilities (15.1% versus 41.3%) (CBS, 2020c).

Empowering Leadership

The variable empowering leadership was measured on a 5-point scale (1 = 'Totally disagree'; 5 = 'Totally agree') using the twelve-item scale developed by Ahearne, Mathieu, and Rapp (2005). An example item is: 'My manager makes many decisions together with me'. The reliability of the scale was good (2019: $\alpha = .93$; 2020: $\alpha = .94$).

4.3.3. Research Design and Statistical Analysis

Our research design resembles a natural experiment in which we exploit the COVID-19 pandemic to estimate the effect of empowering leadership on the well-being of healthcare employees before and during a crisis (Sieweke and Santoni, 2020). To estimate the effect, we apply a DID design. Our approach differs from a classical DID design in a number of ways: we use a continuous treatment variable rather than a binary variable; we do not include an untreated control group; and we analyze within-person changes instead of between-person changes.

In its simplest form (the so-called two groups, two periods design), the DID requires (a) a variable that indicates the period (i.e., before or after the treatment assignment) and (b) a variable that indicates whether a subject was in the treatment or control group. The DID applies a 'double differencing' to estimate the causal effect. That is, we first estimate the difference before and after the treatment assignment both for the control and the treatment group. In the next step, we analyze the difference between these two differences to estimate the treatment effect. Due to this 'double differencing', the DID controls both for time-invariant differences between treatment and control group as well as for time trends that are unrelated to the treatment assignment (Imbens and Wooldridge, 2009).

Our DID differs from the 'classical' DID design in two important ways: First, our study uses a continuous treatment variable, which indicates the intensity of the treatment, instead of a binary variable, which indicates whether a treatment was received or not. Although this approach deviates from most DID studies, we identified several studies that also used a continuous treatment indicator. For instance, Card (1992) used the variation in the fraction of workers who earned less than the newly established minimum wage as treatment intensity measure to analyze the effect of the federal minimum wage on teenagers' labor market outcomes. Similarly, Acemoglu, Autor and Lyle (2004) exploited variation in the mobilization of men for World War II (continuous variable) to investigate the effects of female labor supply on wage structure. Most importantly, some recent econometrics studies indicate that using a continuous treatment indicator is a feasible approach in DID designs. For instance, Angrist and Pischke (2009, p. 234) argued that a 'second advantage of regression DD [difference-in-differences] is that it facilitates the study of policies other than those that can be described by dummy variables,' and other researchers have made similar arguments (e.g., Clair and Cook, 2015; Kahn-Lang and Lang, 2020). Therefore, although the use of a continuous treatment variables deviates from most 'classical' DID studies, it nevertheless represents a valid approach.

Second, our study lacks an untreated control group. This deviation from classical DID studies is due to our research context. Because the COVID-19 pandemic affected all provinces within the Netherlands—although to a varying degree—we could not identify a group of (healthcare) workers who were completely unaffected by the pandemic. Yet even this fact does not threaten the validity of our DID design but rather indicates a different causal interpretation of our estimates. Angrist and Imbens (1995) show, for the instrumental variable design, that models with varying degrees of treatment intensity estimate what they call the 'average causal response', instead of the average treatment effect or the average treatment effect on the treated, commonly estimated in binary treatment models. Callaway and Sant'Anna (2021) extend the work by Angrist and Imbens (1995) to DID designs. A key result of their study is that 'for a continuous/multi-valued treatment, identifying causal response parameters (unlike identifying the treatment effect parameters (…)) does not necessarily require having access to a group that does not participate in the treatment' (Callaway and Sant'Anna, 2021, p. 15). Therefore, our estimates can have a causal interpretation despite the lack of an untreated control group.

Our DID design differs in another way from 'classical' DID studies, which we regard as a considerable advantage of our study: we analyze within-person changes, instead of between-person changes. That is, because we have data from the same employees both before and during the crisis, we use a fixed-effects DID design that compares changes in the level of empowering leadership experienced by employees who work in provinces with low or high exposure to COVID-19. A main advantage of this design is that it controls for all time-invariant employee characteristics (e.g., gender, age etc.). Also, it helps to avoid potential endogeneity that results from leaders adapting their leadership style to fit the needs of their employees (DeRue et al., 2010).

Equation 1 shows our DID model, which we estimated using ordinary least squares (OLS):

$$Y_{it} = \alpha_i + \beta_1 Empowering_{it} + \beta_2 Crisis_t + \beta_3 Crisis_t \, x \, Intensity_j + \beta_4 Crisis_t \, x \, Empowering_{it} + \beta_5 Intensity_j \, x \, Empowering_{it} + \beta_6 Crisis_t \, x \, Intensity_j \, x \, Empowering_{it} + \varepsilon_{it}$$

where Y_{it} is the value of a dependent variable (vigor, dedication, absorption, physical exhaustion and mental exhaustion) of employee *i* in year *t;* α_i are employee fixed effects;

$Empowering_{it}$ represents employee *i*'s empowering leadership score in period *t;* $Crisis_t$ indicates whether employees were surveyed before or during the COVID-19 pandemic; $Intensity_j$ indicates the crisis intensity (i.e., hospital admission per 100,000 inhabitants) within province *j* where the employee worked; and ε_{it} is the error term[1]. We also added several interaction terms to test our hypotheses. We clustered standard errors at the level of the provinces because crisis intensity varies between provinces (Bertrand et al., 2004).

4.4. Results

This section presents the main findings of our study. A confirmatory factor analysis and several additional analyses to check the robustness of our findings and to analyze potential biases are included in Appendix 3.

4.4.1. Correlation matrix

Table 1 presents descriptive statistics and correlations for all variables.

Table 1 *Descriptive statistics and correlations*

				Data from 2019					
Variables	***Mean***	***SD***	**1.**	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**
1. Vigor	3.80	.72	[.84]						
2. Dedication	4.13	.69	0.74	[.82]					
3. Absorption	3.69	.81	0.67	0.77	[.79]				
4. Physical exhaustion	2.40	.83	-0.52	-0.32	-0.31	[.87]			
5. Mental exhaustion	1.82	.75	-0.54	-0.39	-0.27	0.71	[.88]		
6. Crisis intensity	81.22	33.05	0.09	0.03	0.070	-0.09	-0.08		
7. Empowering leadership	3.36	.71	0.31	0.35	0.27	-0.21	-0.16	0.08	[.93]
				Data from 2020					
1. Vigor	3.83	.69	[.84]						
2. Dedication	4.05	.71	0.74	[.82]					
3. Absorption	3.65	.77	0.71	0.75	[.78]				
4. Physical exhaustion	2.26	.81	-0.54	-0.40	-0.33	[.88]			
5. Mental exhaustion	1.94	.78	-0.55	-0.46	-0.33	0.71	[.90]		
6. Crisis intensity	81.22	33.05	0.07	0.07	0.05	-0.12	-0.09		
7. Empowering leadership	3.43	.73	0.35	0.36	0.33	-0.27	-0.26	0.09	[.94]

Note. $n = 468$ in 2019 and 2020; all $r \geq |.09|$ are significant at $p \leq .05$. The variable 'Crisis' was excluded, because of collinearity (i.e., the variable is coded '0' in 2019 and '1' in 2020).

1 Please note that we omitted the direct effect of crisis intensity because it is perfectly correlated with the individual fixed-effects and the interaction terms.

4.4.2. Hypotheses tests

Table 2 presents the results of the three-way interactions of the crisis with empowering leadership and crisis intensity. We analyze the motivational and cognitive dimensions of employee well-being (dedication and absorption, hypothesis ***1a***) and the energetic dimension (vigor, physical exhaustion and mental exhaustion, hypothesis ***1b***).

Table 2 *Fixed-effects (within-person) regressions for the three-way interaction effects of the crisis, crisis intensity and empowering leadership on well-being*

	Model 1 Dedication	**Model 2 Absorption**	**Model 3 Vigor**	**Model 4 Physical exhaustion**	**Model 5 Mental exhaustion**
Crisis	-1.02 (.60)	-.23 (.38)	-1.00[a] (.51)	.54 (.35)	1.23* (.49)
Empowering leadership	.12 (.09)	.32* (.11)	.04 (.16)	-.20[a] (.10)	.08 (.09)
Crisis*crisis intensity	.01 (.008)	-.001 (.004)	.013[a] (.006)	-.008* (.003)	-.01* (.005)
Crisis*empowering leadership	.25 (.17)	.07 (.10)	.31* (.14)	-.18[a] (.10)	-.32* (.14)
Crisis intensity*empowering leadership	.001 (.001)	-.003* (.001)	.001 (.002)	.002 (.001)	-.002* (.001)
Crisis*crisis intensity* empowering leadership	-.003 (.00)	.0002 (.001)	-.004* (.002)	.002[a] (.001)	.003* (.001)
R^2 (within-person)	.08	.03	.06	.07	.07
F	40.95***	5.14**	4.73*	25.74***	38.80***

Note. $n = 936$; unstandardized coefficients are shown (with clustered standard errors in parentheses); [a] $p \leq 0.10$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

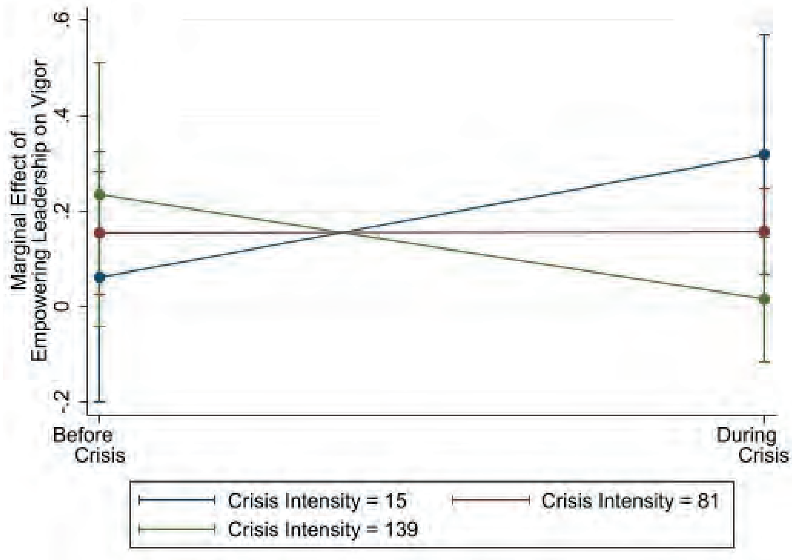
For the motivational and cognitive dimensions, our analysis indicated a non-significant three-way interaction effect regarding employee dedication (Table 2, Model 1; $b = -.003$; $p = .216$) and absorption (Table 2, Model 2, $b = .000$; $p = .833$).

Regarding the energetic dimension, for vigor (Table 2, Model 3), we find a significant and negative three-way interaction effect ($b = -.004$; $p = .048$). Figure 2 shows the three-way interaction plot with the marginal effect of empowering leadership on vigor. The x-axis indicates whether we analyze the marginal effect before (0) or during the crisis (1). For crisis intensity, we use the mean ($M = 81$) and two standard deviations below (M -2 SD = 15[2]) and above the mean (M +2 SD = 139). Figure 2 shows a considerable increase ($\Delta = .26$) in the marginal effect of empowering leadership at low levels of crisis intensity from before the crisis ($\Delta y/\Delta x = .06$; $p = .645$) to during the crisis ($\Delta y/\Delta x = .32$; $p = .013$). At the same time, we find a considerable decrease ($\Delta = -.31$) in the marginal effect of empowering leadership at high levels of crisis intensity from before

2 We used 15 hospitalizations per 100,000 inhabitants (instead of 11), because this is the lowest hospitalization number that was actually observed in our data.

the crisis ($\Delta y/\Delta x = .32$; $p = .013$) to during the crisis ($\Delta y/\Delta x = .01$; $p = .826$). We compared the differences in the changes in the marginal effect of empowering leadership from before to during the crisis at low and high levels of crisis intensity and found that this difference is considerable (.57) and statistically significant ($\chi^2 = 4.95$; $p = .026$).

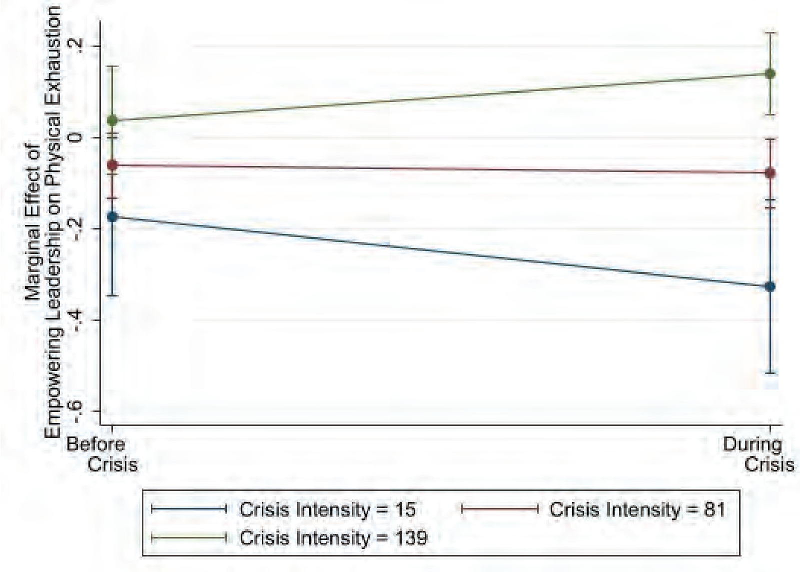
Figure 2 *Moderating effect of the crisis and crisis intensity on the marginal effect of empowering leadership on vigor*



Regarding physical exhaustion, we find a positive three-way interaction effect (Table 2, Model 4; $b = .002$; $p = .059$).[3] Figure 3 shows a considerable increase ($\Delta = -.16$) in the negative marginal effect of empowering leadership at low levels of crisis intensity from before the crisis ($\Delta y/\Delta x = -.17$; $p = .05$) to during the crisis ($\Delta y/\Delta x = -.33$; $p = .001$). We also find an increase ($\Delta = .10$) in the marginal effect of empowering leadership at high levels of crisis intensity from before the crisis ($\Delta y/\Delta x = .04$; $p = .535$) to during the crisis ($\Delta y/\Delta x = .14$; $p = .002$). Again, we compare the differences in the changes in the marginal effect of empowering leadership from before to during the crisis at low and high levels of crisis intensity and find that this difference is considerable ($\Delta = .26$) and statistically significant ($\chi^2 = 4.44$; $p = .035$).

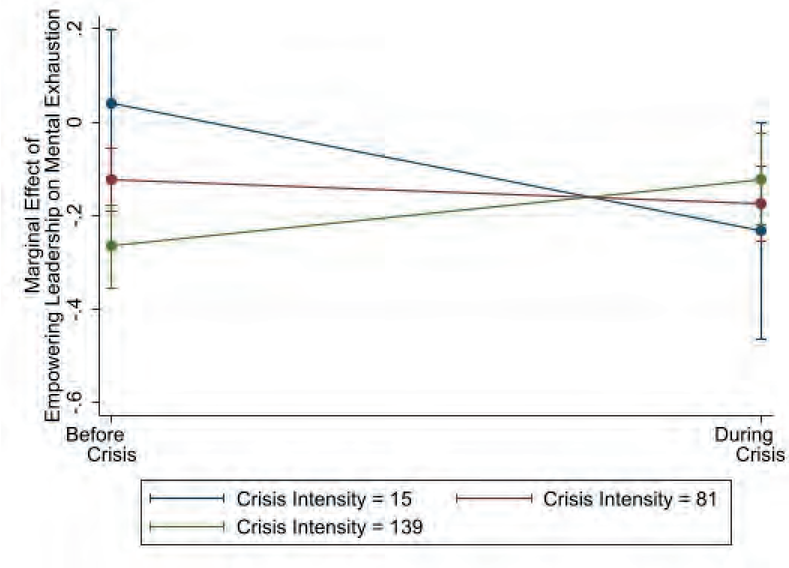
3 Please note that the negative sign indicates a 'positive' effect, because lower levels physical exhaustion indicate a more positive outcome for employees.

Figure 3 *Moderating effect of the crisis and crisis intensity on the marginal effect of empowering leadership on physical exhaustion*



Finally, regarding mental exhaustion (Table 2, Model 5), our analysis indicates a positive and significant three-way interaction effect (b = .003; p = .041). Figure 4 shows, similar to the patterns for vigor and physical exhaustion, that at low levels of crisis intensity the marginal effect of empowering leadership turns negative from before the crisis ($\Delta y/\Delta x$ = .04; p = .621) to during the crisis ($\Delta y/\Delta x$ = -.23; p = .049); this turn is considerable (Δ = -.27). Also, similar to the pattern for vigor and physical exhaustion, we find that at high levels of crisis intensity the negative marginal effect of empowering leadership becomes less negative (Δ = .15) from before the crisis ($\Delta y/\Delta x$ = -.27; p < .001) to during the crisis ($\Delta y/\Delta x$ = -.12; p = .015). A comparison of the differences in the changes in the marginal effect of empowering leadership from before to during the crisis at low and high levels of crisis intensity shows that this difference is considerable (Δ = -.42) and statistically significant (χ^2 = 5.38; p = .02).

Figure 4 *Moderating effect of the crisis and crisis intensity on the marginal effect of empowering leadership on mental exhaustion*



4.5. Discussion

4.5.1. Main findings

In this study, we employed a context-sensitive approach to empowering leadership. Specifically, we investigated whether the effect of empowering leadership on well-being differs during a crisis. To do so, we exploited a public health crisis, the COVID-19 pandemic.

Our findings paint a consistent picture regarding the relationship between empowering leadership and employee well-being during a crisis. First, for the motivational and cognitive dimensions (dedication and absorption) we find that the effect of empowering leadership does not vary during a crisis. Therefore, hypothesis ***1a*** is rejected. Second, for the energetic dimension (vigor, physical exhaustion and mental exhaustion), the effect of empowering leadership on healthcare employee well-being is negatively moderated by a crisis. Therefore, hypothesis ***1b*** is rejected: our study provides clear evidence that empowering leadership becomes less effective during a crisis and may even become a burden and harm employee well-being. These results provide some insights into the effect of empowering leadership on employee well-being during a crisis that we will elaborate on below.

4.5.2. Discussion of our findings and implications

We will now consider how our findings fit the literature on empowering leadership. For one, we are able to address the criticism that studies on empowering leadership show little interest in the role of context (Sims Jr., Faraj and Yun, 2009). What is more, we find that context matters. Contrarily to our expectations based on meta-analytic evidence (Lee, Willis and Tian, 2018; Kim, Beehr and Prewett, 2018), our findings align with those that have a more nuanced view on empowering leadership (e.g., Dennerlein and Kirkman, 2022; Sharma and Kirkman, 2015). Our findings suggest that context, specifically a public health crisis, influences the extent to which empowering leadership is a positive (or even negative) force for employee well-being. Hence, empowering leadership does, again, appear to have two faces (Cheong et al., 2016).

But how do these faces relate, and how does a combination of both lead to consequences for employee well-being? Whilst our understanding is limited, some studies shed more light on these dynamics. Cheong et al. (2016) find that empowering leadership affects performance through a dual process: it increases performance through positive mediation of self-efficacy, but it decreases performance through negative mediation of job-induced tension. In another study, Tuckey, Bakker and Dollard (2012) found empowering leadership increases both cognitive resources as well as cognitive demands that are associated with taking on leadership responsibilities, and that empowering leadership, cognitive resources and demands interact: the combination high empowering leadership, high resources and high demands appeared most optimizing for employee work engagement. Hence, contradictory mediating mechanisms—and their interactions—may explain differential effects of empowering leadership and scholars should take those into account to fully understand the workings of empowering leadership (Cheong et al., 2019). In that sense, a crisis may present such a disruption in employees' balance of demands and resources that employees, being confronted with increasing demands, are not ready for the responsibilities empowering leadership requires (Kniffin et al., 2021; Spoorthy, Pratapa and Mahant, 2020; Bakker et al., 2007). Besides, Demerouti and Bakker (2023) recently extended JD-R theory with a crisis perspective, arguing that in a crisis well-being can be at risk when demands and resources at work interact with those at home or in the organization at large. This could also be why leaders tend to switch to a more directive leadership style after a crisis, a phenomenon known as the threat-rigidity hypothesis (Stoker et al., 2019). In our study we observe a scenario that when a crisis has a relatively low impact, the resources and demands that empowering leadership provides benefit the energetic dimension of employee well-being. When a crisis has a relatively high impact, however, crisis demands hamper potential positive effects of empowering leadership, which diminishes the positive effects on the energetic dimension of employee well-being.

Our results also relate to the study of empowerment in public leadership (Boin and ‘t Hart, 2003). What we find nuances the claim that empowerment is a solution in crises and presents a potential dilemma. From a performance perspective, scholars suggest crises ‘demand lateral coordination, not top-down command and control’ (Boin and ‘t Hart, 2003, p.547), and similarly, empowering strategies may increase employee resilience (McDonald et al., 2016). Yet our results suggest employee empowerment may be challenging in, especially high intensity, crises and may even negatively affect their well-being. From a well-being perspective then, leadership styles that provide direction and confidence to employees are favored as crises get more intense (Antonakis et al., 2003). Simply put, leaders that care about their employees' energy levels should provide direction in crisis situations, rather than dissolving responsibility to lower-level employees.

We should zoom in on the significance of the distinct findings within our operationalization of well-being (Bakker et al., 2014). We found that the motivational and cognitive dimensions respond differently to a crisis compared to the energetic dimension. The three-way interactions are only significant for the energetic dimension (vigor, physical exhaustion and mental exhaustion) and not for the motivational and cognitive dimensions (dedication and absorption). The literature offers a potential explanation. In our theory section we suggested that effects of empowering leadership may be stronger for the energetic dimension of well-being. Especially for the energetic dimension, empowering leadership may be perceived as a necessary resource to counter the increase in demands as this dimension is more prone to a health impairment process that is triggered by stressors than the motivational and cognitive dimensions (Bakker and Demerouti, 2017). Whilst we did not find the expected effects, the fact that we did find effects for the energetic dimension but no effects for the other dimensions may be caused by the same explanatory mechanism. The few studies that found evidence for the dark side of empowering leadership also present negative effects for energy-related variables. For example, Cheong et al. (2016) finds increases in job induced tension through a burdening process, which shows that increasing autonomy may also increase strain (Langred and Moye, 2004). These findings imply that in investigating the effects of empowering leadership, we should be careful not to overgeneralize by measuring dependent variables that are too generic to find these subtle differences.

4.5.3. Practical implications

The practice of context-dependent leadership, which refers to applying the right kind of leadership dependent on what the context requires, is of increasing interest to practitioners (James, 2011). Our study offers valuable practical knowledge as it presents a concrete implication of how situational leadership should be applied. Specifically, we show that leaders and managers, when considering applying empowering leadership in their teams or organizations during a crisis, should consider the intensity of the crisis at hand. This may inform public leaders' behaviors as they attempt to reinvent themselves in response to the challenges they meet (Ansell et al., 2021). In extreme crises, it seems that if leaders want to benefit the energetic dimension of their employees, they should be more directive and less empowering. In mild crises, empowering leadership may very well offer a solution to the challenges the organization faces. We find this to be true, even in a public health crisis as unprecedented as COVID-19.

4.5.4. Limitations and research avenues

Our study has some limitations that require further discussion. First, our study is potentially affected by common source bias, which results from using the same source to collect our independent and dependent variables. Yet, we have at least two reasons to believe that influence of common source bias on our results is negligible. First, our use of within-person analyses significantly reduces potential common source bias because our fixed-effects approach statistically corrects for individuals' response tendencies (Jakobsen and Jensen, 2015), which are an important source of common method bias. Second, studies have shown that it is highly unlikely for common source bias to produce a significant interaction effect in the absence of a true effect (e.g., Siemsen et al., 2010; Lai et al., 2013, p. 243). Therefore, we deem it unlikely that the results of our interaction effects are affected by common source bias.

Second, we should address the extent to which our findings generalize to other crises and samples. Our results may not directly translate to other types of crises. Nevertheless, our results emphasize the inability of empowering leadership to improve well-being when crisis intensity is high. That means that results could be similar in other situations with similar intensity. Whilst the specifics of a crisis may differ, the mechanisms, e.g., high levels of job demands like work pressure rendering empowering leadership ineffective, would be similar (Bakker et al., 2014; Tuckey, Bakker and Dollard, 2012). Future research may address whether this is true. Besides, we study the effects of a crisis on the well-being of healthcare employees. Healthcare employees form a most-likely case, as their work was heavily affected by the crisis (e.g., Greenberg et al., 2020; Zhou et al., 2020). Thus, our study fits the increasing calls for more context-aware research, but more research may be needed to test whether observed results hold for other job sectors.

4.5.5. Conclusion

Empowering leadership is often portrayed as a promising antecedent of employee well-being. Our study provides evidence that the effectiveness of empowering leadership for employee well-being depends on context. Empowering leadership is less effective during a crisis and may even become a burden. Our study on empowering leadership confirms that public leadership in crises, and especially in high intensity crises, is in a league of its own and that it requires a different type of leadership. In these cases, employee well-being may be best protected by leaders who step up to the challenge, take charge and directly deal with the crisis.

Chapter 5

Shared leadership and employee willingness

This chapter is based on the following published article:
Van Roekel, H. (2023). Examining employee willingness to execute shared leadership: The role of leadership behaviour, gender, age, and context. *Leadership, 19*(6), 508-529.

This study is single-authored and therefore written in first-person singular.

Abstract

Shared leadership refers to a post-heroic conceptualization of leadership dispersed among employees. Studies on shared leadership in teams show its emergence depends highly on team and formal team leader characteristics, but employees' own voice is remarkably absent: we know little about how employees individually consider how they would want to execute shared leadership. Taking a bottom-up perspective, this study presents a large-scale conjoint experiment in which 6,742 healthcare employees were asked to evaluate specific leadership behaviours. The results show a notable share of employees are willing to execute shared leadership, but willingness varies dependent on a number of factors. Employees are more willing to share leadership when it is focused on building relationships or bringing about change, when it takes only few hours and when it benefits others. Besides, willingness to execute shared leadership is higher among young or male employees, and in the context of the COVID-19 crisis. This study contributes to understanding how leadership behaviour, personal characteristics and context affect the emergence of shared leadership. The study concludes by critically exploring some of the possible systemic causes for differences in willingness to execute shared leadership, connecting these to broader issues in healthcare employment.

5.1. Introduction

In the past decades, scholars of organizational theory have increasingly questioned the traditional, heroic and hegemonic view of leadership and suggested leadership could be shared by multiple employees (Pearce, Wood and Wassenaar, 2018; Collinson et al., 2018; Tourish, 2015; Pearce, 2004; Pearce and Conger, 2002). Shared leadership can be executed on varying levels in the organization, from inter-organizational levels to inter-individual levels (Ulhøi and Müller, 2014). When shared leadership is executed within teams specifically, 'leadership roles and influence are dispersed among team members' (Zhu et al., 2018, p.836). An increasingly vast literature indicates antecedents of shared leadership are found majorly in the characteristics of the team and formal team leader, and shared leadership is in turn associated with increases in positive team attitudes and team performance (Döös and Wilhelmson, 2021; Wu et al., 2020; Zhu et al., 2018).

Although our understanding of antecedents of shared leadership has increased, we still know little about how employees consider exercising shared leadership on an individual level (Gockel and Werth, 2010), specifically in the public sector (Crosby and Bryson, 2018; Tummers and Knies, 2016). Put differently, employees' own voice and considerations have been somewhat overlooked. This is unfortunate as illustrated by the following thought experiment. Imagine an opportunity for shared leadership arises. What makes employees have the 'willingness and confidence to take on' this leadership role (Pearce and Manz, 2005, p.137)? In the general leadership literature, scholars have developed leadership behaviour taxonomies to study leadership as a collection of specialized behaviours (Yukl, 2002, p.74). Within the shared leadership literature, this approach has rarely been used. Therefore, in this article, I apply insights on leadership behaviour taxonomy to shared leadership, by testing how employees individually evaluate varying specific leadership behaviours (Yukl, 2002; Jønsson et al., 2016). Hence, the present study questions: which shared leadership behaviours do employees want to exercise?

In a conjoint experiment on Dutch healthcare employees, I conceptualized shared leadership into multiple leadership behaviours that employees are supposed to exercise, also including the effort that these behaviours require, and the beneficiary or beneficiaries that these behaviours are likely to have. Besides, I assessed whether individual willingness varies dependent on employees' personal characteristics, specifically gender and age. Finally, during the design of the study, the COVID-19 crisis hit and severely affected healthcare systems (Zhou et al., 2020). Due to its unprecedented consequences, I also added this as a relevant contextual variable to the conjoint design.

This study contributes to the shared leadership literature by finding how shared leadership may or may not emerge on an individual level dependent on (a) the specific shared leadership behaviour, (b) the person that is supposed to exercise shared leadership, and (c) the context they find themselves in. Hence, this study looked beyond the higher-order concepts that shared leadership usually focuses on, like the general level of shared leadership (Carson et al., 2007; Avolio et al., 2003), and beyond the more often studied shared leadership antecedents on a team level (Döös and Wilhelmson, 2021; Wu et al., 2020; Zhu et al., 2018). In doing so, this study opens up the conceptual black box of shared leadership and studies it like leadership itself is often studied: as a collection of specialized behaviours, executed by an individual (Yukl, 2002). Importantly, this study includes, and aims to further foster, a critical discussion of the systemic issues that underlie individual differences in willingness, like socio-cultural norms in organizations and differences in the romanticization of leadership between men and women. Besides, this approach allows the consideration of the promise as well as the potential pitfalls that shared leadership presents. Finally, this study responds to pleas to study shared forms of leadership in public settings (Crosby and Bryson, 2018; Tummers and Knies, 2016). There is also a practical contribution: if (HR) managers want to explain current levels of shared leadership in organizations or if they want to successfully implement shared leadership, this study shows how they can take into account what specific leadership behaviours employees are more willing to execute.

To address the research question, I conducted a conjoint experiment (Hainmueller et al., 2015; 2014). Conjoint experiments were developed in mathematical psychology (Luce and Tukey, 1964), have been often used in marketing and consumer research to assess consumer preferences (Green and Srinivasan, 1990) and, more recently, in political science to study voter preferences (Hainmueller et al., 2014). Very few studies on leadership have applied conjoint methods (exceptions include Tavares et al., 2018). Within shared leadership studies specifically, most empirical studies rely on surveys and case studies (Ulhøi and Müller, 2014). Conjoint analysis offers a valuable addition to the literature's methodological toolbox as, among its many benefits, it allows the assessment of individual preferences regarding shared leadership on a granular level (Wu et al., 2020). Besides, a conjoint experiment mitigates social desirability bias and multicollinearity (Horiuchi et al., 2022; Karren and Barringer, 2002) and is shown to more closely represent real-world behaviour (Hainmueller et al., 2015).

5.2. Theory

5.2.1. Shared Leadership

Traditionally, leadership is defined as the process of social influence towards certain goals exercised by people with formal leadership roles in the organization towards their followers (Antonakis and Day, 2017). Yet, already a century ago, Mary Parker Follet argued leadership should originate from the individual with the best skills in a particular situation (Follet, 1924; Fitzsimons et al., 2011). This claim marks an explicit questioning of the traditional, hegemonic view of leadership as a heroic, individual quest (Tourish, 2015; Collinson et al., 2018). Notions like these preceded the conceptualization of shared leadership, which centres on one question: could leadership be something that many members of a team share (Pearce and Conger, 2002; Pearce, 2004)? The study of shared leadership has since then spread over social science disciplines and has been explored in numerous public and private contexts (e.g., Pearce, Wood and Wassenaar, 2018; Crosby and Bryson, 2018; Ulhøi and Müller, 2014). Not surprisingly then, many definitions of shared leadership exist (Wu et al., 2020; Zhu et al., 2018; Ulhøi and Müller, 2014). Carson et al. (2007, p.1221) point out shared leadership 'can take place in a team with or without a designated leader, can be either formal or informal, and addresses the distribution and sharing of leadership among all team members, in contrast to only one or two leaders'. Zhu et al. (2018) argue that most definitions share three similarities, regarding the source of leadership, the unit of analysis and the distribution. Shared leadership, in their view, can be understood as a) lateral influence amongst peers, b) that can emerge within a team, and c) is characterized by a dispersion of roles and influence among team members (p.836).

There are multiple concepts that are related to shared leadership. Most closely related are concepts like distributed leadership (Barry, 1991). Like shared leadership, distributed leadership has been studied in a variety of academic disciplines and contexts (Tian et al., 2016). It has been used to study, for example, self-managed teams within the disciplines of organizational behaviour and HRM (Barry, 1991), network structures of leadership perceptions in the private sector (Mehra et al., 2006), and leadership practices in schools within educational research (Fitzsimons et al., 2011). More recently, Jønsson et al. (2016) developed a scale to measure distributed leadership within healthcare contexts. Besides distributed leadership, collective leadership is often used interchangeably with shared leadership (e.g., Avolio et al., 2009). A related but slightly different topic concerns collaborative leadership: sharing leadership can happen at varying levels of collaboration (Sullivan et al., 2012). Subtle differences between the concepts remain open to debate (e.g., Currie and Lockett, 2011).

5.2.2. Antecedents of Shared Leadership

Ulhøi and Müller (2014) find relatively few studies address the antecedents of shared leadership. Nevertheless, Zhu et al. (2018), Wu et al. (2020) and Döös and Wilhelmson (2021) present integrative frameworks linking shared leadership to its antecedents and outcomes.

Most shared leadership antecedents that have been studied originate in either factors relating to characteristics of the team or the formal team leader (Döös and Wilhelmson, 2021; Wu et al., 2020; Zhu et al., 2018). For example, shared leadership is facilitated by an internal team environment that allows employees to share a purpose, express mutual social support, and have a voice (Carson et al., 2007, pp.1222-1223). Also, team member characteristics, like integrity and trust are positively related to shared leadership (Drescher et al., 2014). Besides, transformational and empowering leadership of leaders (Hoch, 2013), and leader humility (Chiu, Owens and Tesluk, 2016) stimulate shared leadership.

5.2.3. Conceptualizing Shared Leadership

Although many antecedents have been studied, the literature is notably silent about the willingness of employees to execute shared leadership (Gockel and Werth, 2010). There is a viable reason for this gap in the literature related to its conceptualization. Shared leadership has been conceptualized mainly in one of two ways. Some studies approach shared leadership as a generic concept. For example, Carson et al. (2007, p.1225) measure the amount of general leadership behaviour by asking 'to what extent does your team rely on this individual for leadership?'. Paletta (2012) measures distributed leadership in schools by measuring the ratio of leaders to teachers. Molenveld et al. (2021) distinguish conveners, mediators and catalysts as different types of shared leaders. Other studies measure how a specific leadership style is shared by adapting an existing leadership scale. For example, Avolio et al. (2003) measure shared transformational and transactional leadership.

These two approaches are still relatively abstract and focus on leadership as a generic process. In contrast, Yukl (2002) presents a way of conceptualizing leadership that has been largely overlooked in the shared leadership literature. He argues leadership is a collection of specialized behaviours, such as networking, problem solving and encouraging innovation (Yukl, 2012, p.74). A notable exception is Jønsson et al. (2016), whose distributed leadership scale is based on an earlier version of Yukl's leadership behaviour taxonomy. One of the few mentions of this behaviour-focused shared leadership approach is voiced by Pearce et al. (2008, p.626), who argue shared leadership theory is an 'explicit attempt' at 'integrating the view of leadership as a role performed by an individual with the view of leadership as a social process'. Similarly, Ospina (2017, p.280) argues 'the primary source of leadership is not the person but the role that he or she takes up'.

If leadership is a collection of specialized behaviours, shared leadership is the allocation of one (or more) of these behaviours to an employee in a non-leadership position. Conceptualizing shared leadership as the sharing of a specific leadership behaviour has the advantage that it is possible to assess how the willingness of employees alters according to the type of shared leadership behaviour. Imagine that in a team there is a certain, hitherto formal, leadership behaviour that presents an opportunity for shared leadership. This means that employees have the choice to share this leadership behaviour and will consider whether they have 'the willingness and confidence to take on part of the leadership role' (Pearce and Manz, 2005, p.137). It is likely that their consideration on whether to exercise shared leadership will include what exactly it is that they would have to share. In the literature on shared leadership there has been little attention to factors that affect employees' willingness. Therefore, below I develop a framework to test the assumption that shared leadership behaviour matters for willingness to execute shared leadership (Gockel and Werth, 2010).

In Table 1, I conceptualize shared leadership behaviour. Shared leadership requires exercising a specific behaviour (e.g., the behaviour is representing the team at board meetings), which requires a specific level of effort from an employee (e.g., the behaviour costs two hours per week for a period of three months), and includes a beneficiary of beneficiaries (e.g., the behaviour benefits especially colleagues). These features present by no means an exhaustive list but they are a coherent set of features based on the literature. Together, these three features will enable exploration of factors that influence employees' willingness to execute shared leadership, and also allow to analyze how they differ across personal characteristics and context.

Table 1 *Three features of shared leadership*

Feature	Working definition
Behaviour	The specific shared leadership behaviour someone has to exercise
Effort	How much time the shared leadership behaviour requires and whether this can be substituted
Beneficiary	The target group that is most likely to benefit from the shared leadership behaviour

5.2.4. Shared Leadership Behaviour

First, employees will consider the leadership behaviour they are required to exercise. Table 2 presents the leadership behaviour taxonomy by Yukl (2012), who distinguishes between four main dimensions of leadership behaviour. The first three—task, relation and change—describe internal team leadership, whereas external behaviour refers to leadership on behalf of and outside the team. It is likely that the latter will be less favored among employees, because shared leadership is an emergent *team* phenomenon. Research shows that vital antecedents for shared leadership to emerge within the team include warmth, trust and peer support (e.g., Carson et al., 2007; Drescher et al., 2014; Zhu et al., 2018). These antecedents emphasize the importance of a safe and familiar environment, whose emergence in the team does not guarantee a similar environment outside of it. What is more, the extant leadership literature confirms that employees are generally less conscious of external leadership behaviour because they have fewer opportunities to see their leaders exercise this leadership (Yukl, 2012). Therefore, employees may be less drawn to external behaviours.

Table 2 *Leadership behaviour taxonomy (cf. Yukl, 2012, pp. 68-74)*

Dimension	Definition	Examples
Task	'To ensure that people, equipment, and other resources are used in an efficient way to accomplish the mission of a group or organization.'	Planning, Monitoring
Relation	'To enhance member skills, the leader-member relationship, identification with the work unit or organization, and commitment to the mission.'	Supporting, Developing
Change	'To increase innovation, collective learning, and adaptation to external change.'	Advocating change, Encouraging innovation
External	To 'provide relevant information about outside events, get necessary resources and assistance, and promote the reputation and interests of the work unit'.	Networking, Representing

Besides, whereas behaviours focused on task or relation behaviour are meant to maintain the status quo, e.g., providing necessary resources and caring for team members, a change behaviour is meant to challenge existing situations by, e.g., encouraging innovation (Yukl, 2012). Pearce (2004) argues that two of the characteristics that especially call for shared leadership concern tasks that are high in complexity and tasks that require a lot of creativity (see also Fitzgerald et al., 2013). The need for shared leadership may also be increased when the organization faces turbulent times (Lund and Andersen, 2023), is struggling to achieve its goals (Günzel-Jensen et al., 2018). In contrast, for routine tasks, 'the need for any type of leadership (…) is minimal' (Pearce, 2004, p.50). The emergence of shared leadership is closely linked to the observation that 'today's employees desire more from work than just a paycheck; they want to make a meaningful impact' (Pearce, 2004, p.47). As change behaviours are more complex, require more creativity and have more impact, employees are more drawn to change leadership behaviours.

H1a: *Employees are less willing to exercise external leadership behaviour compared to other leadership behaviours.*

H1b: *Employees are more willing to exercise leadership behaviour aimed at change compared to other internal leadership behaviours.*

5.2.5. Effort

Second, employees will consider the effort a shared leadership behaviour requires. Specifically, whilst employees may be willing to exercise a shared leadership behaviour, they are likely to be less attracted to extra workload. Employees have only limited work time to employ and increases in workload have shown to be related to, e.g., more burnout (Bakker and Demerouti, 2017) and absenteeism (Van Woerkom et al., 2016). Therefore, employees are likely to be drawn more to shared leadership behaviours with less additional workload. Another avenue through which executing a shared leadership behaviour does not imply higher workload, is if in exchange for it, the amount of regular work is decreased. Indeed, Pearce argues that in order for shared leadership to succeed, the preconditions include 'securing necessary resources' (2004, p.51). If employees, in exchange for taking on a shared leadership behaviour, are provided with sufficient resources, they may be more likely to exercise such a behaviour.

The hypotheses express effort as the sum of intensity (i.e., how many hours a week), longevity (i.e., how many months), and subsidiarity (on top or in exchange for regular work):

H2a: *Employees are more willing to exercise leadership behaviour that take fewer hours per week compared to more hours per week.*

H2b: *Employees are more willing to exercise leadership behaviour that has shorter longevity compared to longer longevity.*

H2c: *Employees are more willing to exercise leadership behaviour when exchanged for regular work compared to when exercised on top of regular work.*

5.2.6. Beneficiary

Third, employees will consider the beneficiary or beneficiaries that the shared leadership behaviour is likely to have. Shared leadership is about making meaningful impact (Pearce, 2004). This impact is primarily directed towards others, as shared leadership is argued to improve group-level caring through increases in psychological empowerment climate and group solidarity (Houghton et al., 2015). In other words: 'sharing is caring' (Houghton et al., 2015, p.313). Similarly, research shows endorsing collectivistic views will more likely lead to shared leadership (Hiller et al., 2006). Therefore, employees are more drawn towards behaviours that are likely to benefit others rather than themselves.

Finally, research shows 'shared leadership has benefits for work teams beyond just improving team processes' (Carson et al., 2007, p.1229). Many studies show shared leadership affects the 'ultimate' goal: increases in team performance, like higher client or customer satisfaction (Zhu et al., 2018; Houghton et al., 2015). In most organizations, performance is still a huge factor in formal reward systems, and in the end, employees will engage more in behaviours that are expected from them (Pearce, 2004, p.51; D'Innocenzo, Mathieu and Kukenberger, 2016). Therefore, employees are more drawn to shared leadership behaviours that are likely to directly affect performance measures, meaning they will value benefits for clients (e.g., patients) over benefits for peers.

H3a: *Employees are more willing to exercise leadership behaviour that benefits others compared to behaviour that benefits themselves.*

H3b: *Employees are more willing to exercise leadership behaviour that benefits clients compared to behaviour that benefits peers.*

5.3. Methods

To assess the above, I executed a conjoint experiment. In this conjoint experiment, respondents were asked to choose between two shared leadership scenarios. I applied a conjoint experiment with prompted choice, meaning respondents were forced to choose. Conjoint experiments with prompted choice are shown to satisfactorily mimic real choice (Hainmueller et al., 2015): when respondents have to choose between alternatives, they will take more effort evaluating them. Both scenarios include a number of attributes (i.e., the conjoint term for variables, e.g., which specific leadership behaviour) and levels (i.e., all possible values for a variable, e.g., a task leadership behaviour). For each respondent, each of the two scenarios presents a random level for each attribute and across the two scenarios attributes are presented in a random order to counter survey order effects (Krosnick and Alwin, 1987). By analyzing respondents' choices, a conjoint experiment allows the assessment of multiple hypotheses at the same time (Hainmueller et al., 2015; Hainmueller et al., 2014).

The main benefit of using a conjoint experiment in the context of shared leadership is that it allows individual preferences of employees to be addressed. Most of the literature on shared leadership has used aggregating measures or social network analysis, hereby largely ignoring individual considerations (Wu et al., 2018). Besides, conjoint experiments offer multiple methodological benefits above more common survey methods in shared leadership research (Ulhøi and Müller, 2014). First, conjoint experiments mitigate social desirability bias because respondents' attention is drawn away from the most sensitive attributes (Horiuchi et al., 2022). Therefore, estimates about employees' preferences regarding shared leadership may be more accurate than if they were assessed with regular survey measures. Second, surveys can suffer from multicollinearity when independent variables correlate. In contrast, conjoint methods allow the researcher to experimentally manipulate variables to truly assess their independent effects (Karren and Barringer, 2002). Third, conjoint experiments more accurately represent real-world behaviour because respondents, like in real life, make multidimensional choices (Hainmueller et al., 2015; Karren and Barringer, 2002).

I set up the conjoint experiment for Qualtrics using the Conjoint Survey Design Tool (Strezhnev et al., 2019). This tool, together with a webserver, allowed the loading of the conjoint experiment as a web element into the survey.

5.3.1. Survey and Sample

This experiment was included in a large annual survey. Data collection was approved by the Faculty Ethical Review Committee of Utrecht University. Before analyzing the data, the research description, hypotheses and methods were preregistered at the Open Science Framework (See List of preregistrations). The survey was sent out via a Dutch healthcare employee collective in May-June 2020. At becoming a member of this collective, respondents agreed to be sent emails, including the one that invited them to participate in this survey. Before participating in the survey, respondents provided their informed consent. Following standard procedure, I presented respondents with information about the goal of the survey, the procedure of participation (including opting out after participation in the survey, this was possible until four weeks after the first survey invitation), data storage, processing and usage, and ways to get in touch with the researcher. After reading this information, respondents gave their active informed consent by clicking 'Yes, I agree' to being informed by the goal of the survey, being able to ask any questions, understanding the way the data are treated, and understanding the opt out procedure. For those that chose 'No, I don't agree', the survey was terminated.

This study did not use deception: no intentional misleading was used and information relevant to the study was provided beforehand (see section 'Experimental setup'). The only information that was withheld was that the survey questions were about leadership rather than generic additional tasks, but this is very unlikely to pose any potential risk to respondents. Therefore, no debriefing was provided directly after the survey experiment. However, at the end of the general survey, respondents were given instructions on how to receive the survey results and interpretation in their inbox.

A total of 6,742 respondents participated in the experiment (each respondent evaluated two vignettes at the same time, so observations for the analyses are double the number of respondents). Respondents are Dutch healthcare employees who indicated they do not primarily fulfil formal leadership positions. Table 3 presents respondent characteristics. The sample is representative for the Dutch healthcare sector in terms of gender (in the Dutch healthcare sector, 84.3% is female), yet less representative in terms of age (in the Dutch healthcare sector, 34% is younger than 35, and 24.2% is older than 55 years). The respondents represent all healthcare industries, with an overrepresentation of hospitals and an underrepresentation of mental health care (Dutch healthcare percentages: hospitals: 21.9%; nursing homes and homecare: 32.1%; mental health care: 7%; disability care: 13.4%; other: 4.7%) (CBS, 2020a).

Table 3 *Respondent characteristics (n = 6,742)*

	Percent
Gender	
Female	83.4
Male	16.6
Age	
Mean (*SD*)	51.4 (10.14)
< 25	1
26-35	8.0
36-45	18.4
46-55	30.0
56-65	41.5
66 >	1.1
Healthcare industry	
Hospitals	36.5
Nursing homes and homecare	23.4
Mental health care	16.7
Disability care	16.7
Other	6.7

5.3.2. Attributes

Table 4 presents the levels for the included attributes. Operationalization has been executed combining the theoretical insights presented above with practical insights from healthcare professionals. The professionals informed the research by suggesting realistic levels for the attributes. For example, they suggested maintaining a maximum of 16 hours a week for the attribute on intensity, and they advised on realistic tasks (under 'behaviour') that would resonate with healthcare professionals. At the same time, these behaviours had to maintain the notion of the general leadership behaviour.

During the design of this study, the COVID-19 crisis severely affected healthcare systems around the world (Zhou et al., 2020). This crisis increased job demands among many healthcare workers, especially affecting the well-being of employees working with COVID-19 patients (Van Roekel et al., 2021). Consequentially, effects of COVID-19 on the willingness to take on challenges like shared leadership are likely as well (Kniffin et al., 2021). Because of this, and following an increasing call that leadership studies address the important societal issues (Tourish, 2017), I specifically address this contextual variable in this study's design. In the conjoint experiment, I assessed whether it matters if the leadership behaviour is supposed to be exercised during the COVID-19 crisis, or afterwards.

Table 4 *Operationalization of shared leadership*

Attribute ► Level ▼	Shared leadership behaviour	Effort of shared leadership behaviour: Intensity	Effort of shared leadership behaviour: Longevity	Effort of shared leadership behaviour: Subsidiarity	Beneficiary of shared leadership behaviour	Crisis (control)
1	Finding ways to decrease administrative burden within the team (Change)	2 hours a week	A month	On top of regular tasks	Your self	During COVID-19 crisis
2	Chairing the team meetings (Task)	4 hours a week	A quarter	In exchange for less regular tasks	Your peers	After COVID-19 crisis
3	Becoming counsellor for the team (Relation)	8 hours a week	A year	-	Your patients	-
4	Representing the team at the board meetings (External)	16 hours a week	Entire employment	-	-	-

5.3.3. Experimental Setup

Upon digitally entering the experimental part of the survey, I presented respondents with two vignettes (translated in Dutch) displaying all of the attributes in random order, and randomly presenting a level for each attribute. Table 5 presents an example. I asked them to choose between the two scenarios via the following accompanying message: 'This question is about executing additional roles at your job. Below, there are two descriptions of additional roles you could exercise. Please point out which role you would be more interested in: role 1 or role 2. You can only choose one role and you must choose'. After respondents chose one of the vignettes, I asked them to evaluate both: 'What is the chance you would exercise this role if you were asked to?' Respondents could answer on a five-point Likert scale ranging from 'very low to 'very high'. I did not specifically refer to the concept of leadership as I simply wanted them to evaluate a role I defined as leadership behaviour, rather than confounding that evaluation with respondents' general interpretation of leadership. Also, note that this design focuses on the individualistic side of shared leadership, whether a respondent wants to exercise a shared leadership behaviour or not, rather than shared leadership as a team dynamic.

Table 5 *Example scenario*

	Shared leadership behaviour A	**Shared leadership behaviour B**
The role will be	finding ways to decrease the administrative burden within the team	chairing the team meetings
The role will benefit	your peers	your patients
The role will cost you	2 hours a week	8 hours a week
In total, the role will last for a period of	a month	a year
You will exercise this role	on top of your regular tasks	in exchange for less regular tasks
Starting	after the COVID-19 crisis	during the COVID-19 crisis

5.3.4. Analysis

After data collection, I transformed and analyzed the data in R, using the Cjoint package (Hainmueller et al., 2014). My main analysis consisted of calculating the Average Marginal Component Effect for every attribute (AMCE). The AMCE is the 'marginal effect of each randomized attribute, averaged over the joint distribution of all attributes' (Jilke and Tummers, 2018, p.234). Following Hainmueller et al. (2014), I calculated the AMCE using a linear regression model with employees' behaviour choice as the outcome variable. Besides, to assess whether results differ over gender and age, I conducted identical analyses over subsets of the data (cf. Hainmueller et al., 2014). Again, note that the analyses present sample sizes that are double the number of respondents, as each respondent evaluated two vignettes at the same time, choosing one and rejecting another. Finally, I reported the results of the evaluation of each behaviour of the respondents, in total and for the subsets[1].

5.4. Results

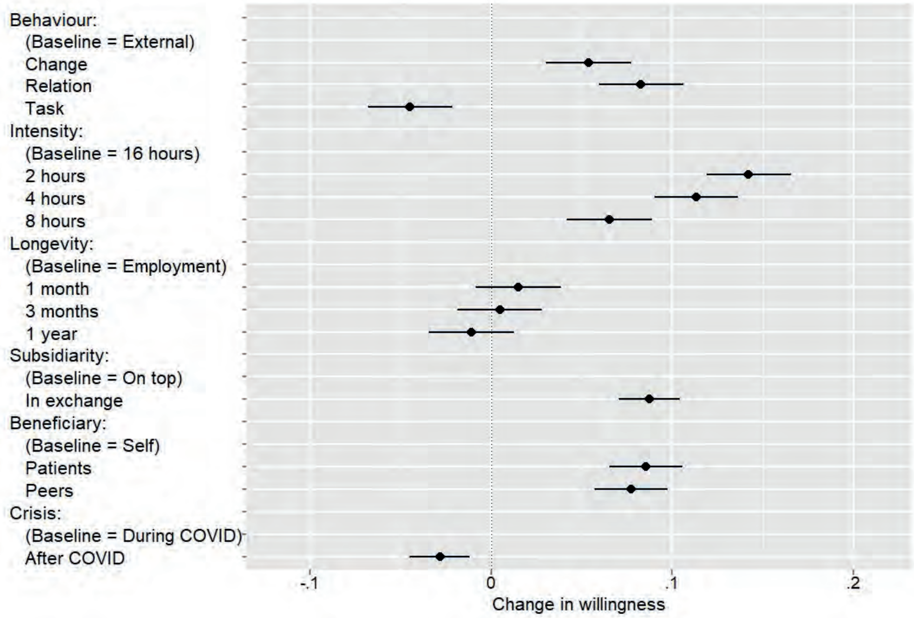
5.4.1. Main Analysis

I present the results of the analysis in Figure 1 and Table 6. In the figure, for each conjoint attribute level, regression coefficients are represented with the dot, extended with 95% confidence intervals. Hence, when confidence intervals do not cross the zero

1 The main data and syntax for this study are accessible at: https://osf.io/ekdr6/?view_only=f802587d0ff-3434cabc1a9a80c42a7eb .

line, it shows that the specific coefficient is statistically significant compared to the baseline of the attribute (at 95% confidence level). This means that the specific conjoint attribute level, which is a leadership behaviour feature, is associated with a significantly higher or lower probability of willingness to exercise leadership behaviour compared to the attribute level set as baseline.

Figure 1 *Results of main analysis: Average Marginal Component Effect of all attributes on change in willingness. (95% confidence intervals, n = 13,484)*



First, for behaviour, I find that, compared to external leadership behaviour (leading the external connections of the team), employees are 5% ($p < .001$) more likely to choose change behaviour (leading innovation, learning and change within the team), 8% ($p < .001$) more likely to choose relation behaviour (leading the improvement of intra-team relations), but 4% ($p < .001$) less likely to choose task behaviour (leading in accomplishing tasks efficiently within the team). This means that hypothesis ***1a*** is partially accepted: employees are not more willing to execute external behaviour than change and relation behaviour but they are more willing to execute external behaviour than task behaviour. Also, hypothesis ***1b*** is partially accepted: employees are more willing to execute change behaviour than task behaviour, but they are not more willing to execute change behaviour than relation behaviour.

Table 6 *Results of main analysis. *** p < .001*

Attribute	Baseline	Level	Estimate (SD)
Behaviour	External	Change	.05*** (.01)
		Relation	.08*** (.01)
		Task	-.04*** (.01)
Intensity	16 hours	2 hours	.14*** (.01)
		4 hours	.11*** (.01)
		8 hours	.07*** (.01)
Longevity	Employment	1 month	.01 (.01)
		3 months	.005 (.01)
		1 year	-.01 (.01)
Subsidiarity	On top	In exchange	.09*** (.008)
Beneficiary	Self	Patients	.09*** (.01)
		Peers	.08*** (.01)
Crisis	During COVID-19	After COVID-19	-.03*** (.008)
n	13,484		

Second, for effort, I find that, compared to behaviour that costs 16 hours per week, employees are 7% ($p < .001$) more likely to choose behaviour that costs 8 hours, 11% ($p < .001$) more likely to choose behaviour that costs 4 hours, and 14% ($p < .001$) more likely to choose behaviour that costs 2 hours. In contrast, no significant differences are found for longevity. Finally, employees are 9% ($p < .001$) more likely to choose behaviour when in exchange for work rather than on top of it. This means that hypotheses ***2a*** and ***2c*** are accepted: employees are more willing to execute behaviours that take fewer hours than behaviours that take more hours, and they are more willing to execute behaviours when exchanged for regular work compared to on top of regular work. Hypothesis ***2b*** is rejected: longevity does not matter.

Third, for beneficiary, compared to behaviour that benefits themselves, I find that employees are 8% ($p < .001$) more likely to choose behaviour that benefits peers and 9% ($p < .001$) more likely to choose behaviour that benefits patients. This means that hypothesis ***3a*** is accepted: employees are more willing to execute behaviours that benefit others than behaviours that benefit themselves. In contrast, hypothesis ***3b*** is rejected: employees are not more willing to execute behaviours that benefit patients than behaviours that benefit peers.

Finally, I find an effect for crisis: employees are 3% ($p < .001$) less likely to choose behaviour to be carried out after the COVID-19 pandemic compared to during the pandemic.

5.4.2. Context Analysis

To test whether employee characteristics matter, I carried out identical conjoint experiments on subsets of the data and checked whether there were notable differences. These differences can only be interpreted as differing preferences relative to baseline categories, and not as descriptive, absolute differences in preference (Leeper et al., 2020).

First, I compared the subset of male employees (n = 2,240) versus the subset of female employees (n = 11,244). Since the majority of the main sample is female, the results for the female sample are similar to the main sample. However, for the male sample, there are differences. Male employees are not more willing to choose change and relation behaviours compared to the baseline of external behaviour, they are not more willing to choose behaviours benefitting patients compared to behaviours benefitting themselves, and they are not less willing to exercise behaviours after the crisis compared to during the crisis.

Second, I compared the subset of younger employees (younger than the mean age, < 51.4, n = 5,746) versus the subset of older employees (older than or equal to the mean age, >= 51.4, n = 7,738). Younger employees (mean age = 41.5, SD = 7.2) are not less willing to choose a task behaviour compared to the baseline of external behaviour, and older employees (mean age = 58.7, SD = 3.9) are not less willing to exercise behaviours after the crisis compared to during the crisis.

5.4.3. Chance Analysis

Finally, employees indicated the chance they would exercise the presented behaviours if they were asked to. Table 7 presents the mean willingness for the sample and the subsamples, separately for the left and right vignette presented to respondents. First, there is a notable share of employees willing to exercise leadership behaviour: 3 (medium chance) and 4 (high chance) are the most chosen scores. Second, there are no significant differences between the left and right vignette, indicating that biases due to vignette position are unlikely. Third, there are significant differences between the subsamples. For vignette one, I find male employees to be significantly (p < .05) more willing to than female employees. For both vignettes, I find younger employees to be significantly (p < .001) more willing to than older employees.

Table 7 *Means, standard deviations and t-tests for the chance employees would exercise specific leadership behaviours on a five-point Likert scale ranging from 'very low to 'very high'*

	Mean (*SD*)		Total *n*
	Left vignette	**Right vignette**	
Main sample	2.85 (1.17)	2.87 (1.21)	13,484
Male employees	2.92^{*a} (1.20)	2.92 (1.22)	2,240
Female employees	2.84^{*a} (1.17)	2.86 (1.21)	11,244
Younger employees	3.03^{***b} (1.13)	3.02^{***c} (1.19)	5,746
Older employees	2.72^{***b} (1.19)	2.76^{***c} (1.22)	7,738

Note. T-tests with 95% confidence intervals are executed to calculate whether significant differences exist a) between the two vignettes and b) between the subsamples (e.g., young vs old). *** $p < .001$. * $p < .05$. The superscript letters indicate which means are significantly different from each other. The sample size is equally split among the two vignettes as each respondent evaluates two vignettes.

5.5. Discussion

5.5.1. Main Findings

5

The central aim of the present study was to investigate which shared leadership behaviours employees want to exercise. In a two-scenario conjoint experiment with prompted choice, I randomly varied the leadership behaviour employees are supposed to display, the effort the leadership behaviour requires, and the beneficiary or beneficiaries the leadership behaviour is likely to have. Besides, I studied whether willingness of employees varied based on gender and age, and dependent on the COVID-19 context.

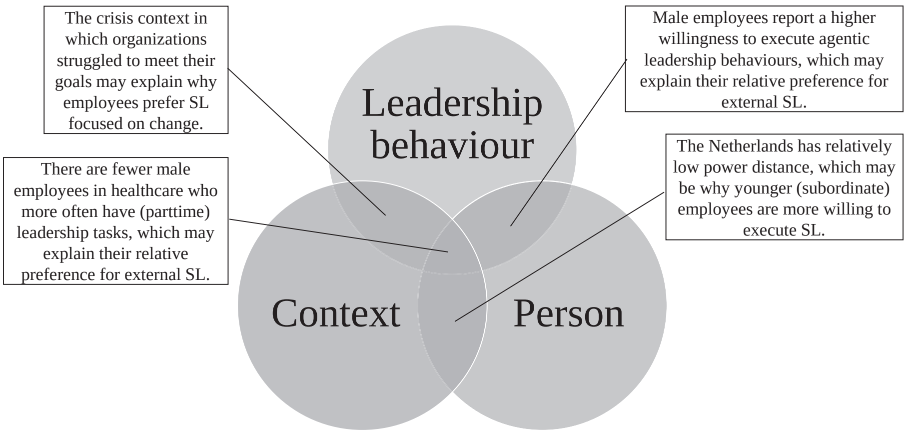
The results show that employees are more willing to execute relation and change leadership behaviours than external leadership behaviours, but they are still more willing to execute external leadership behaviours than task leadership behaviours. They are also more willing to execute shared leadership behaviours that take fewer hours per week than behaviours that take more hours, but longevity does not matter. And they are more willing to execute behaviours that benefit clients or peers than behaviours that benefit themselves. There are some differences in willingness between employees based on gender and age. The most striking difference is that the relative dislike of external behaviours does not seem apparent for male employees. Besides, a large share of employees is willing to exercise the leadership behaviours as presented, but most notably male and younger employees more so. Finally, employees are more willing to exercise shared leadership in a COVID-19 context.

5.5.2. Implications and Contributions

This study has employed a new approach to assessing the emergence of shared leadership, by conceptualizing leadership as a collection of specialized behaviours, as is common in the literature on leadership behaviour taxonomy (Yukl, 2002), but rarely used within the context of shared leadership. Rather than investigating on a team level, individual employees were questioned about their willingness. As a result, this study improves our understanding of the emergence of shared leadership on an individual level, beyond general antecedents on a team level (Döös and Wilhelmson, 2021; Wu et al., 2020; Zhu et al., 2018).

Figure 2 visualizes the three key contributions. This study finds that shared leadership emerges on an individual level dependent on the leadership behaviour an employee has to execute, the characteristics of the person involved and the context they are in. The main findings have described these factors separately, yet the factors may also have combined explanatory power, of which examples are given in the figure. Below, I elaborate on the three key contributions, their combined explanations, and the extent to which the results match the hypotheses and existing literature.

Figure 2 *Shared leadership emerges on an individual level based on leadership behaviour, person and context*



*Note***.** Examples shown are potential explanations of the results using combinations of factors. SL = shared leadership.

First, shared leadership emerges dependent on the specific leadership behaviour. In this study, I distinguished specific behaviours based on the leadership behaviour taxonomy, as well as the effort and beneficiary of those behaviours. The first contribution of this study is that employees are more willing to execute shared leadership behaviours focused on change and relations than other behaviours. This suggests that not all leadership behaviours are equally suitable for shared leadership. While slightly differently than hypothesized, a likely explanation for the higher willingness regarding relation behaviours is that the quality of relationships between employees in a team is shown to motivate shared leadership behaviours and may thus especially motivate behaviours that, in turn, foster relationships (Carson et al., 2007; Drescher et al., 2014; Zhu et al., 2018). The higher willingness regarding change behaviours is in line with earlier findings in the general leadership literature that there is more meaningful impact in leadership behaviours that are complex and require creativity, something that may have been exacerbated by the challenges organizations faced during COVID-19 (Lund and Andersen, 2023; Yukl, 2012; Pearce, 2004). In contrast, the fact that for routine tasks the need for leadership is little and less meaningful (Pearce, 2004) could be a cause of the relative dislike for task behaviours, behaviours that are concerned with leading in accomplishing tasks efficiently within the team (Yukl, 2012). Also, external leadership may be less popular as shared leadership tends to emerge within teams and employees have fewer opportunities to observe external leadership, something that may be exacerbated even more by the hierarchical nature of the healthcare sector (Zhu et al., 2018; Yukl, 2012).

Another implication of this study is that employees want shared leadership that is manageable: they are more willing to execute behaviours that take less effort compared to more effort and in exchange for other work, as too much increase in workload may have negative consequences for employee well-being (e.g., Bakker and Demerouti, 2017). In contrast, a potential explanation for the lack of effects on longevity is that people notoriously underestimate how their preferences will change over time, known as the presentism bias (e.g., Bauckham et al., 2019).

When it comes to benefits from shared leadership, this study suggests that shared leadership is primarily driven by collectivistic rather than individualistic reasoning (Houghton et al., 2015; Hiller et al., 2006). However, the difference between clients and peers is small, which suggests the proposed mechanism that employees engage especially in those behaviours that are rewarded (i.e., behaviours benefitting clients which improves performance, Pearce, 2004), does not apply here. Peer support may be as big of a driver of shared leadership as is client care, as employees want to be generally prosocial (Carson et al., 2007).

Second, shared leadership emerges dependent on the person. This study finds male employees are generally more willing to exercise shared leadership, and that male employees do not share the relative dislike of external behaviours with female employees. This might show a parallel with a recent meta-analysis on the gender gap in general leadership, which found that men are still more likely to emerge as a leader. Men report a higher willingness to execute agentic leadership behaviours (Badura et al., 2018). The findings of this study may imply that the gender gap often observed in general leadership is also present in shared leadership. There are multiple factors to be considered here.

First, it is important to consider the differences in conceptualization of leadership between leadership emergence and shared leadership. Leadership emergence is understood as the degree to which an individual is perceived by others as a leader (Badura et al., 2018), whereas this study uses employees' own assessments of willingness. At the same time, multiple studies have found men to report higher self-estimates of their leadership abilities than women (see Fleenor et al., 2010). Another difference is that shared leadership, by definition, implies multiple people can be a leader, whereas for leadership emergence approaches vary. Some studies analyze leader elections or rankings (with one 'winner'), while others employ scale ratings per employee. The latter approach is more compatible with shared leadership and is shown to be a significant methodological moderator between gender and leadership emergence (Badura et al., 2018).

Second, the particular healthcare sector context affects this study's findings. Research indicates that healthcare leaders are often male (Mousa et al., 2021). A 2019 World Health Organization report, titled '*Delivered by women, led by men*', confirmed that gender-based gaps in leadership are present across the global healthcare workforce: women represent 70% of the workforce but hold only 25% of the senior positions. While none of the employees in the sample primarily fulfilled a leadership position (this was an eligibility criterium), male employees (16.6% in the sample) may feel more invited to leadership. In organizations affected by sexist norms, men may be more socialized than women to feel a willingness to display shared leadership behaviours.

Third, this study addressed individual preferences. However, such preferences are related to larger and systemic issues, not empirically addressed in this study, which play a role when it comes to gender and leadership (Ryan et al., 2016; Fleenor et al., 2010). It is crucial to critically discuss these systemic issues and their consequences, as failing to do so would essentialize gender differences instead of approaching them as products of socialized norms. In this section I aim to start this critical discussion, and I also refer to important literature on the topic to foster further discussion. The aforementioned WHO report (2019) points out how the gender leadership gap within healthcare is perpetuated by, among other things, stereotypes (e.g., the socialized norm that equates women to nursing positions but not managerial roles), sexism (e.g., not appointing women to leadership positions because of sexism, or appointing women for different leadership tasks compared to men) and deep-rooted hierarchies within patriarchal organizations (see also Ryan et al., 2016). Hence, systemic issues like sexism foster a discriminatory environment in which female employees are treated differently and do not get the same opportunities as male employees (Ryan et al., 2016). As a consequence, willingness to execute shared leadership may depend on how leadership is perceived differently by men and women. Within the critical feminist literature, scholars have shown how notions of heroically taking up leadership positions are often tied to masculinity (Collinson, 2011; Elliott and Stead, 2008). Scholars have suggested future research should study whether men romanticize leadership to a greater extent than women (Collinson et al., 2018). The question that this study poses is whether shared leadership, although defined as a post-heroic form of leadership, may still perpetuate gender differences. If so, shared leadership could still be dominated by men who, because of an environment with systemic inequality, may feel more drawn to exercise leadership (compare this to Khan et al.'s (2022) notion of post-heroic heroes or Schweiger et al.'s (2020) sense of self-as-a-leader). Finally, explanations of gender effects may also be found beyond the workplace. Studies show COVID-19 has exacerbated the unequal division of care burdens between women and men (Power, 2020; Carli, 2020), which may further hamper female employees' opportunities to participate in additional work activities like executing shared leadership behaviours.

Besides gender, I find age effects: younger employees indicated significantly higher willingness to exercise shared leadership than older employees. This finding contributes to a literature with mixed findings (compare e.g., Chaturvedi et al., 2012). A potential explanation for the finding is that for older employees, there is less motivation to do more than required and work their way up the career ladder (Zacher, Rosing and Frese, 2011). Likewise, in the context of the Netherlands, a country with relatively low power distance, younger employees may take more initiative even if they are more often subordinate employees (Hofstede, Hofstede, and Minkov, 2010). As the sample is older than the population, all analyses on the (relatively) younger and older subsamples should be interpreted with some caution. However, analyzing the younger subsample separately also provides relevant information for the very reason that the sample is older than the population, as a way to control for age effects.

Third, I found the COVID-19 crisis had an effect on the emergence of shared leadership. As the COVID-19 crisis struck healthcare systems around the world, calls for leadership intensified (Kniffin et al., 2021). Following calls for leadership studies to address such societal changes (Tourish, 2017), I consider that several mechanisms may play a role here. A potential mechanism is that the crisis increased the need for and meaningfulness of shared leadership (e.g., Lund and Andersen, 2023), causing healthcare employees to favor shared leadership during a crisis (Pearce, 2004). However, a slight variation of presentism bias—'we are in a crisis now'—may also have played a role (e.g., Bauckham et al., 2019).

5.5.3. Limitations and Future Research

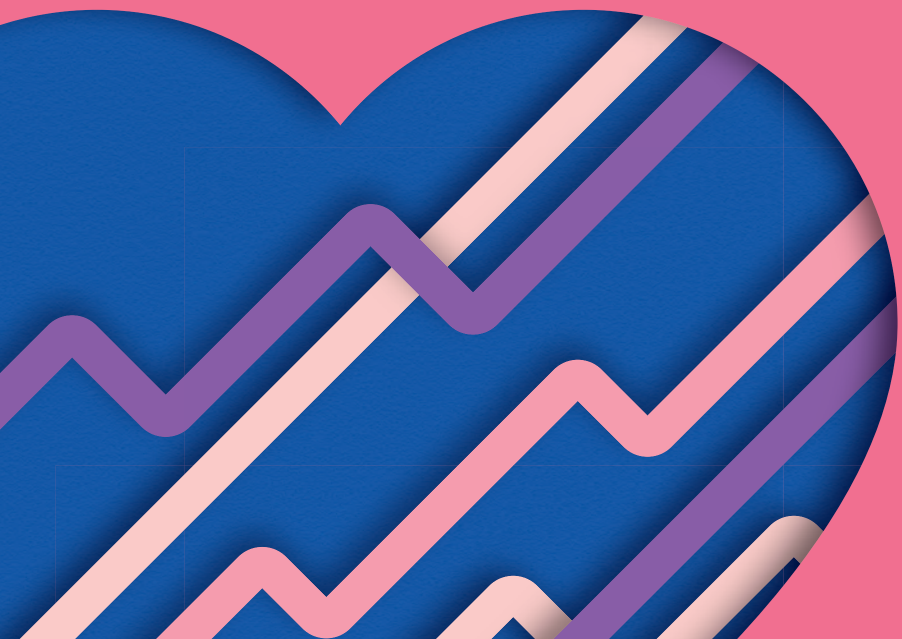
The main limitation, inherent in much survey research, is that I measure self-reported preferences rather than actual preferences or behaviour. As a result, a gap between intent and behaviour may exist that can be explained by other mechanisms. For example, respondents may be subject to social desirability bias, which suggests respondents may consider some answers as socially desirable and this may inflate findings (Grimmelikhuijsen et al., 2017). Nevertheless, researchers argue social desirability bias is likely less problematic in this study as leadership is not a very sensitive, personal or private matter (Lee, Benoit-Bryan and Johnson, 2012). What is more, design choices also helped to combat social desirability bias: respondents were anonymous and not under direct observation of the researcher. Finally, in this conjoint experiment employees were forced to choose between leadership behaviours that differed on multiple, randomly presented attributes. Recent research suggests that in doing so, conjoint experiments may counter social desirability bias because respondents' attention is drawn away from the most sensitive attributes (Horiuchi et al., 2022).

Furthermore, limitations apply to the way the study is designed. It is important to emphasize that preferences in a conjoint experiment should always be interpreted as relative to the presented baseline categories (e.g., employees are more willing to execute change leadership than external leadership) rather than descriptively (e.g., employees like change leadership) (Leeper et al., 2020). Furthermore, while differences in preferences are significant in this large-scale sample, some differences that I found are small. Besides, in this study, three elements (behaviour, effort and beneficiary) were assessed, and they were operationalized in a specific way. For example, I used a taxonomy of leadership behaviours to define shared leadership behaviours (Yukl, 2012). This assumption could be questioned: shared leadership behaviours and leadership behaviours may not necessarily overlap entirely. Nevertheless, many influential studies within the shared leadership literature have successfully deduced shared leadership behaviours from formal leadership behaviours (e.g., Carson et al., 2007; Avolio et al., 2003). More fundamentally, this study focuses on shared leadership behaviours as a role that employees are supposed to execute. This resembles a more formal approach to shared leadership. However, Carson et al. (2007) have already stated shared leadership can manifest itself in both formal and informal ways. By focusing on a more formal role, this study has not addressed ways in which shared leadership emerges informally and how employees take up leadership behaviours as informal activities (Zhu et al., 2018). Finally, this study has not addressed the extent to which willingness coincides with competencies (i.e., knowledge and skills) needed to execute shared leadership. This study has addressed willingness (i.e., employees' attitude), yet whether employees who are willing are also capable of sharing leadership is another question (a point closely related to found discrepancies between self- and other-ratings of leadership, see for example Fleenor et al., 2010).

Overall, this study provided evidence of the factors that affect the willingness of employees to execute formal shared leadership roles within teams. Due to its specific assumptions and operationalization, it has necessarily neglected many other factors and their interactions: how willingness relates to competencies (i.e., knowledge and skills) or opportunities (i.e., contextual factors), how shared leadership is executed in informal activities rather than formal roles, how shared leadership relates to other forms of collective leadership, and how shared leadership is executed on different levels than the team, to name a few. In future research, these topics as well as other avenues may be studied. For example, qualitative studies may inductively explore what conceptualizations employees come up with themselves and dive deeper into employees' own understanding of shared leadership. Through practices of workers inquiry, we could acquire more immersive evidence on how leadership is shared (Smolović Jones et al., 2022).

5.5.4. Conclusion

Using insights from the general leadership literature, this study aimed to contribute to understanding the emergence of shared leadership among employees in teams. With a conjoint experiment, this study has shed light on the willingness of employees to execute shared leadership. This study found three factors that are important to employees' individual considerations about shared leadership: the specific leadership *behaviour* an employee has to execute, including the effort hereof and the beneficiary, the characteristics of the *person* involved and the *context* they are in. Ultimately, this study has modelled itself after the spirit of shared leadership: shared leadership implies giving a voice to individual team members, and research on shared leadership should do the same.



Chapter 6

Motivation activation to increase ethical reporting

This chapter is based on the following published article:
Van Roekel, H., & Schott, C. (2022). Activating employees' motivation to increase intentions to report wrongdoings: evidence from a large-scale survey experiment. *Public Management Review*. DOI: 10.1080/14719037.2021.2015184.

Abstract

Public servants are frequently confronted with unethical behaviour. Research shows intentions to report wrongdoings are increased by activating public service motivation (PSM). We study whether public servants display different reactions to different wrongdoers and whether intentions are also affected by prosocial motivation (PM). We employed a survey experiment on 11,728 healthcare workers. The results show activating PSM or PM increase intentions to report patients, but not colleagues. However, effects are small. What is more, activation of PM has a larger effect for respondents with lower PM-levels. We discuss implications for the literature on the interplay between ethics and motivations.

6.1. Introduction

'Adherence to the highest standard of ethical conduct is inherent to the mission of government organizations' (Belle and Cantarelli, 2017, 327). Nevertheless, across the world public servants frequently engage in, or are confronted with, unethical behaviour at the workplace, such as attempts to obscure the truth, rule bending or even breaking, and outright lying (e.g., Meyer-Sahling, Schuster, and Mikkelsen, 2018; OPM (US Office of Personnel Management), 2012). In addition, public servants are often confronted with verbal aggression (e.g., calling names and rudeness) or sometimes even physical aggression from clients (e.g., Baron and Neuman, 1998; Tummers et al., 2016). This raises the question of how public organizations can deal with these forms of wrongdoing conducted by employees and clients.

Meyer-Sahling, Mikkelsen, and Schuster (2019) found that the activation of public service motivation (PSM), which refers to an 'individual's orientation to delivering services to people with a purpose to do good for others and society' (Perry and Hondeghem, 2008, vii), increases the intentions to report wrongdoings. This finding is highly interesting as it suggests that public organizations aiming to improve ethical behaviour may benefit from activating their employees' PSM. However, the literature on whistleblowing suggests that the reporter-wrongdoer relationship is an important factor for the intentions to engage in reporting (e.g., King and Hermodson, 2000; Treviño and Victor, 1992; Zipparo, 1999). Inspired by this stream of literature, we believe that it is important to go further and pay closer attention to the question on whom employees report. As the context of this study is the health sector, we focus on colleagues and patients as two categories of wrongdoers.

In addition, besides PSM, other motivations and motives are regularly studied in public management literature (e.g., Pedersen, Andersen, and Thomsen, 2020; Breaugh, Ritz, and Alfes, 2018; Lapworth, James, and Wylie, 2018), and they may play an important role in advancing ethical decision-making. Specifically the effect of prosocial motivation (PM), i.e., 'the desire to expend effort to benefit other people' (Grant, 2008a, 48), could be considered as it presents a second type of other-oriented motivation increasingly studied in the field of public administration (e.g., Van der Voet, Steijn, and Kuipers, 2017; Ritz et al., 2020), frequently being associated with prosocial behaviour (Bolino and Grant, 2016).

Motivated by these two research avenues, this study aims to answer the following question: Does the activation of PSM and PM influence the intentions to report wrongdoings from colleagues differently compared to the intentions to report wrongdoings from patients?

An answer to these questions is highly relevant as it not only contributes to the limited research on (un)ethical behaviour in the field of public administration in general (Belle and Cantarelli, 2017) and the PSM-ethics relationship in particular (Ripoll, 2019), but also to the discussion on the distinctiveness of motivational concepts studied in public management research (Ritz et al., 2020; Schott et al., 2019). It does so by studying (1) the causal effect of two types of other-regarding motivation (PSM and PM) on intentions to report wrongdoings as potential factors for improvement, and (2) testing whether the causal relationship between PSM/PM and intentions to report wrongdoings depends on the question who committed the misconduct (i.e., colleagues versus patients).

6.2. Theory

6.2.1. Introducing ethical behaviour and intentions to report wrongdoings

The concept of moral or ethical behaviour refers to a wide range of behaviours and can broadly be defined as behaviour that is subject to (or judged and evaluated according to) generally accepted norms of behaviour (Reynolds and Cerani, 2007; Trevino, Weaver and Reynolds, 2006). This definition closely aligns with James Rest's (1986) classic four-component model of ethical behaviour, which has also been applied within the public management literature (c.f., Loyens and Measschalck, 2010; Ripoll and Breaugh, 2019). Central to this model is the idea that ethical behaviour is the product of four subsequential steps connected by feedback and feedforward loops: awareness, judgement, intention, and actual behaviour. Following recent research by Meyer-Sahling, Mikkelsen, and Schuster (2019), we focus on the third step of Rest's model and study public servants' intentions of to report wrongdoings or ethical problems to management. In the context of (health-)care, wrongdoings refer to violations of important professional and organizational norms, thereby providing a threat to the well-being of employees, patients, and the organization (Searle and Rice, 2020). Because norms cannot be fully captured by official structures and processes, wrongdoings entail both formal and informal overstepping of boundaries.

The intentions to report wrongdoings have, amongst other things, been studied in the literature on whistleblowing, which can be defined as 'the disclosure by organization members (former or current) of illegal, immoral, or illegitimate practices under the control of their employers, to person or organizations that may be able to effect action' (Near and Miceli, 1985, 4). The theoretical foundation of whistleblowing draws from socio-psychological theories based on prosocial behaviour and bystander intervention

(Wise, 2000). When an employee encounters an unethical or illegal act, the person has the three options: (1) exiting the context, (2) ignoring the act, or (3) taking action and blowing the whistle (Miceli and Near, 1992). Following Pillay, Reddy, and Morgan (2017), who study whistleblowing intentions, we assume that 'the choice of behaviour will result from one's intentions' (p.426). This assumption is not only in line with Ajzen's (1991) classic Theory of Planned Behaviour, but also more recent evidence from meta-analyses suggesting that ethical intentions are closely associated with actual ethical behaviour (Armitage and Conner, 2001; Hertz and Krettenauer, 2016). However, we also need to be aware of the fact that for behaviour related to whistleblowing, intentions and behaviour may diverge (Mesmer-Magnus and Viswesvaran, 2005). We come back to this issue in the discussion section of this article.

In addition, based on insights from whistleblowing literature, we need to acknowledge that the intentions to report wrongdoings and actual reporting may not always align automatically with ethical behaviour. Some scholars argue that those who (intent to) blow the whistle do this out of moral concerns for the well-being of others (O'Sullivan and Ngau, 2014; Watts and Buckley, 2017). From this perspective, reporting wrongdoings presents an in-role behaviour in that public servants are legally and morally bound to maintain high standards of integrity and concern for the public interest (e.g., Lavena, 2016; Taylor, 2018). However, there are also scholars suggesting that whistleblowers are 'malcontents' (Devine and Aplin, 1986) who are being driven by selfish concerns such as personal vengeance. For example, Andon and colleagues (2018) found that whistleblowing rates increase when people receive monetary or personal gain for blowing the whistle.

Inspired by the literature on whistleblowing we also argue that it is important to consider the question of—who is the wrongdoer in the conduct of misbehaviour?—when studying ethical reporting. In the whistleblowing literature, the reporting of wrongdoings is typically categorized in terms of whether the reporter and the alleged wrongdoer are peers or not (De Graaf, 2010). Peer reporting is often considered a specific type of whistleblowing (Treviño and Victor, 1992). Other wrongdoers are the employing organization, with management as the responsible party, and clients or customers of the reporter. This differentiation is important because the reasons to report have been argued to be straightforward in all categories except for peer reporting (De Graaf, 2010). For example, one consistent finding is that the severity of the wrongdoing increases the likelihood of an organizational member to step up (e.g., Miceli and Near, 1992; Miethe and Rothschild, 1994). However, the loyalty towards direct colleagues is often much stronger than loyalty towards the institutions or organizations of which one is a member (Pershing, 2003). A reason for this is the frequent contact with direct colleagues, as loyalty has been found to arise out of daily interactions (Sztompka, 1999). Many public service employees have a long history with their direct colleagues whom they see much more frequently than (higher) managers, or customers and clients they are responsible for. Circumstances facilitating strong loyalties towards colleagues are closed systems, such as highly professionalized groups of employees or the policing force, where loyalty and devoted allegiances to direct colleagues present attributes of utmost importance (Heck, 1992). Consequentially, in the medical world, under-reporting of errors made by colleagues presents a serious problem (e.g., Lawton and Parker, 2002; Scott and Henneman, 2017). Based on this, we argue that it is important to specify the wrongdoer when studying the intentions of public servants to report wrongdoings to management.

Given the importance of ethical behaviour in general and the importance of reporting intentions for managing the integrity of the public service (Taylor, 2018), the question of how it can be affected is highly relevant. In the next section two motivational concepts are introduced that potentially affect intentions to report wrongdoings: public service motivation and prosocial motivation.

6.2.2. Introducing public service motivation and prosocial motivation

In the Public Management literature the concept of public service motivation (PSM) has often been used to explain other-regarding behaviours of public servants (Ritz, Brewer, and Neumann, 2016) and it is commonly defined as 'an individual's orientation to delivering services to people with a purpose to do good for others and society' (Perry and Hondeghem, 2008, vii). Typically, values, norms, and beliefs within public institutions can be seen as the origin of PSM (Perry and Vandenabeele, 2008).

Prosocial motivation (PM) is another type of other-regarding motivation, which can be defined as 'the desire to expend effort to benefit other people' (Grant, 2008a, 48). This type of motivation originates from the field of psychology but is increasingly discussed and studied in Public Management literature (e.g., Van der Voet, Steijn, and Kuipers, 2017; Ritz et al., 2020).

The definitions of PSM and PM illustrate the similarity of the concepts: both are primarily directed at others rather than the self. However, there are also differences. 'Instead of being actuated by the wish to explain behaviour such as altruism and self-sacrifice, as has been the case in PSM research, research on PM has been stimulated by questions such as how employees can be motivated to 'care about contributing to other people and the organization' (Grant, 2009, 94)' (Schott et al., 2019, 1).

This raises the question of whether PSM and PM advance intentions to report wrongdoings in the same way. The next section elaborates on the relationship between PSM and reporting intentions as well as the relationship between PM and reporting intentions, while also paying attention to the wrongdoer of misbehaviour (patient versus colleague).

6.2.3. Public service motivation and the reporting of wrongdoings

Research addressing the relationship between PSM and ethics in general (e.g., Choi, 2004; Kwon, 2014; Maesschalck et al., 2008) and PSM and whistleblowing (e.g., Brewer and Selden, 1998; Caillier, 2015) relays on a straightforward argument. It is commonly assumed that there is a positive effect of PSM on ethical outcomes because PSM and ethics share the same public values, which can be described as 'the desire or need to act in ways that promote the public interest' (Wright, Hassan, and Park, 2016). This argument has recently been refined to explain counter-intuitive approaches and findings from research on PSM and ethics (Schott and Ritz, 2018; Christensen and Wright, 2018). In particular, Ripoll (2019) builds on the work of scholars viewing PSM as public service identity grounded in public institutions (e.g., Bednarczuk, 2018; Schott et al., 2015; Perry and Vandenabeele, 2008) that nourish individuals' PSM by transmitting their institutional logics and values through different social processes such as socialization and social learning (Perry and Vandenabeele, 2008). This makes PSM a self-concept imbued with public values. Highly public service-motivated individuals are expected to consistently regulate their (un)ethical decisions and behaviours in line with the set of values, norms and rules that shaped their public service identity (Ripoll, 2019). In addition, identity theory (Stets and Burke, 2003) helps to explain PSM produces ethical behaviour. A central idea of this theory is that individual engage in specific behaviours as they want to signal to others and themselves who they are: a process called self-verification (Stets and Burke, 2003; Schott and Ritz, 2017). With regard to public service identity, this means that the more strongly they are committed to this identity, the more likely they are to engage in ethical activities as this this type of behaviour reflects what they find important. When differentiating between patients and peers as wrongdoers and considering the question whether public service motivated individuals are equally willing to report wrongdoing conducted by both groups, we argue the answer is likely to be 'yes' for the following reasons.

Blowing the whistle on peers, which occurs 'when group members go outside their group to report a member's misconduct' (Trevino and Victor, 1992, 30) does not only threaten the group's authority structure and reputation, but can also have severe consequences for the whistle-blower her- or himself. Loyalty and solidarity among staff members lead them to perceive the whistle-blower as a betrayer and excluding the person from the group (Loyens, 2013). While much whistleblowing literature is about organizational wrongdoing (De Graaf, 2010), less is known about patient wrongdoing and subsequent blowing the whistle on patients. For this reason, we use insights from the inspection literature (Van der Walle and Raaphorst, 2018) to learn more about the consequences of whistleblowing when the wrongdoer is the reporter's client. For example, Schott (2015) found that inspectors' decision not to intervene and to overlook misbehaviour from their inspectees was influenced by the desire to maintain good working relationships and the feared consequences for the person who engaged in the misbehaviour. This suggests that patient reporting—just like peer reporting—does not only have negative consequences for the wrongdoer, but also for the reporter.

When turning towards PSM and the question of how this type of motivation helps to stimulate peer and patient reporting, most scholars would agree that PSM 'goes beyond self-interest and organizational interest' and motivates 'individuals to act accordingly whenever appropriate' (Vandenabeele, 2007, 549). Based on this, we argue that the question of who the wrongdoer is may be irrelevant for individuals with high levels of PSM. Rather, PSM seems to be a 'general motivation' to do good (Schott et al., 2019, 1203), which is unsensitive to potential conflicts with the interests of others and one's own interests.

Following the successful approach of Meyer-Sahling et al. (2019, 450), we aim to causally assess our core expectations on the effects of PSM on intentions to report wrongdoings through randomly activating the concept. The central premise is that asking respondents about a certain concept will activate that concept and render it salient in that moment. The theoretical argument for the effect of activation is based on social identity theory (Meyer-Sahling, Mikkelsen, and Schuster, 2019). This theory suggests that individuals have multiple identities that are activated (or not) depending on the situation they are in. By randomizing whether respondents are asked questions before or after outcome variables, we vary the situations that respondents are in and can assess the causal effect of this activation (Meyer-Sahling, Mikkelsen, and Schuster, 2019, 450). This approach takes inspiration from experiments with low-intensity activation of PSM (Pedersen, 2015), experiments that use micro-interventions to increase awareness of prosocial and societal impact (Vogel and Willems, 2020), as well as from the idea that survey questions can function as prime and affect answers to following questions (Zaller and Feldman, 1992). The question order priming method has been applied across disciplines. Insights include that bank employees behave more dishonestly when their professional identity is activated (Cohn, Fehr and Marechal, 2014). And activating the criminal identity of inmates from a medium security prison has been found to increase cheating behaviour (Cohn, Maréchal, and Noll, 2015). Within public administration, studies have shown the existence of these primes in citizen satisfaction (Andersen and Hjortskov, 2016; Van de Walle and Van Ryzin, 2011) and user satisfaction (Thau et al., 2021).

Importantly, the findings from Meyer-Sahling, Mikkelsen, and Schuster (2019) are not uncontested. Other studies relying on student samples did not find evidence that PSM activation leads to actual and/or self-reported ethical behaviour (Christensen and Wright, 2018; Olsen et al., 2019). One potential explanation for the mixed findings may be differences in sample characteristics.

Lastly, Meyer-Sahling, Mikkelsen, and Schuster (2019) hypothesized the effect of activation may be larger for respondents with higher levels of PSM because there simply is more PSM to activate. This argument may seem to neglect the possibility of ceiling effects, making it dissimilar to Linos (2018), who suggests a PSM treatment is less effective for those with more public service orientation in the first place. Yet, Linos (2018) aims to present new information to future public sector workers, whilst Meyer-Sahling, Mikkelsen, and Schuster (2019) are merely rendering this information salient among a large sample of Chilean central government employees, which is why larger effects are expected for those with higher levels of PSM.

We put the above arguments to the test by our first pair of hypotheses:

H1: *Activating PSM increases intentions to report wrongdoings from patients.*

H2: *Activating PSM increases intentions to report wrongdoings from colleagues.*

H3: *Activating PSM has a larger effect on intentions to report wrongdoings from patients and colleagues for respondents with higher levels of PSM.*

6.2.4. Prosocial motivation and the reporting of wrongdoings

To our knowledge, there is no research investigating the direct relationship between PM and intentions to report wrongdoings. However, there is a large body of research demonstrating the effect of PM on helping behaviours (see, for example, Bolino and Grant (2016) for a literature study) and the likelihood to take initiative (De Dreu and Nauta, 2009). Reporting wrongdoings can be considered as helping others who are affected by misbehaviour and as taking initiative, suggesting a link between PM and intentions to report wrongdoings as well. Again, identity theory helps to explain this relationship. Individual with a prosocial identity—i.e., individuals scoring high on PM—are likely to engage in these behaviours because they want to signal to themselves and others who they are: people who care about others and who want to change things for the better.

However, when turning to the question whether prosocially motivated individuals are equally willing to report wrongdoing conducted by colleagues no clear answer can be given. Remember, PM—or 'the desire to expend effort to benefit other people' (Grant, 2008a, 48)—can be directed towards different beneficiaries such as individuals, groups or large collectives such as organization (Grant and Berg, 2011). The tricky thing about stepping up and reporting a wrongdoing is that its positive consequences for one party always come along with cost for one or more other parties. Reporting a wrongdoing from a colleague probably means trouble not only for the co-worker who engaged in misbehaviour, but also for the professional group one belongs to. At the same time, the quality of the services may improve, benefitting the receivers of services such as patients. When considering the intentions to report a wrongdoing from a patient, we see a similar pattern. While the quality of services is likely to improve for other patients just like the working conditions of colleagues, the patient him- or herself is likely to suffer from being outed as wrongdoer.

This means reporting or not presents prosocially motivated individuals with a dilemma. As their PM can be directed towards different beneficiaries—individuals (i.e., the wrongdoer) or groups (i.e., group of colleagues or group of patients) (Grant and Berg, 2011)—we cannot predict how the person will decide. After all, reporting misbehaviour from a peer may benefit the group of patients but hurt the misbehaving colleague and/or professional group one belongs to. Reporting misbehaviour from patients may benefit a group of patients and/or professional group one belongs to, but makes the life of the wrongdoer (i.e., the misbehaving patient) more difficult. Depending on which beneficiary PM is directed at, the intentions to report can be expected to either increase or decrease.

Indirect support for this line of reasoning is provided by Bolino and Grant (2016), who argue that 'when the intention is to benefit the group or the organization, some [prosocially motivated] employees may [even] engage in unethical prosocial behaviour. These actions may take form of commissions like criticizing other workgroups in order to enhance their own's team standing, or omissions such as withholding negative information about the organizations' products or services' (618). Similarly, Somers and Casal (1994) argue that loyalty towards beneficiaries can interfere with recognizing and reporting violations of justice and ethics. Again, we aim to causally assess our core expectations on the effects of PM on intentions to report wrongdoings through randomly activating the concept. This leads to our second set of hypotheses.

H4-A: *Activating PM increases intentions to report wrongdoings from patients.*

H4-B: *Activating PM decreases intentions to report wrongdoings from patients.*

H5-A: *Activating PM increases intentions to report wrongdoings from colleagues.*

H5-B: *Activating PM decreases intentions to report wrongdoings from colleagues.*

H6: *Activating PM has a larger effect on intentions to report wrongdoings from patients and colleagues for respondents with higher levels of PM.*

6.3. Materials and methods

6.3.1. Context

The survey experiment was embedded in a survey on work and health of Dutch healthcare employees that was issued by IZZ, a large healthcare employee collective in The Netherlands[1]. Only employees currently working in a healthcare organization were included. The survey was sent via Qualtrics to approximately 144,692 respondents, who are members of IZZ. The survey was conducted in March and April 2019. In total, 12,260 respondents participated (response rate = 8.47%), of which 11,728 completed the survey experiment.

Table 1 shows the percentages for gender, age and branch of the respondents besides the same statistics for the Dutch healthcare employees that are a member of IZZ. Some differences existed. Most noticeable is that the 'other'-category for branch was underrepresented in the sample. Furthermore, our sample was somewhat older and included more females. Whilst generalization is to some extent limited, we conclude our sample moderately represents the population of healthcare employees that are member of IZZ.

1 This study involves human participants and was reviewed and approved by the Faculty Ethical Review Committee of Utrecht University. The participants provided their written informed consent to participate in this study.

Table 1 *Percentages sample versus population*

	Sample (12,260)*	IZZ healthcare employees**
Gender (F/M)		
Female	81.6%	74%
Male	18.2%	26%
Age (18-64)		
18-29	4.6%	8%
30-44	22.8%	26%
45-64	72.6%	66%
Branch		
Hospitals	37.1%	37%
Nursing/Care/Home care	23.8%	23%
Mental healthcare	16.9%	14%
Disabled care	16.5%	15%
Other	5.7%	11%

Note. * We use our initial sample. ** The descriptives are obtained from IZZ. No absolute numbers are provided due to competition sensitivity.

6.3.2. Measures

To measure the independent variables (PSM and PM) Dutch translations of two four-item scales were used (Vandenabeele and Jager, 2020; Grant, 2008b). We used a unidimensional measure of PSM as we were interested in the core concepts of PSM. Whilst both uni- and multidimensional scales have been developed for PSM (Kim et al., 2013), Vandenabeele, Ritz, and Neumann (2018) argue global measures enable to focus on the main driver of the concept. PM is traditionally measured by unidimensional scales (e.g., Grant, 2008a, 2008b).

To measure the dependent variables (intentions to report wrongdoings from colleagues, or patients) two single items were used that are adapted from Meyer-Sahling, Mikkelsen, and Schuster (2019). These items were altered to add the wrongdoer. Meyer-Sahling, Mikkelsen, and Schuster (2019) argue using a measure of ethical behavioural intentions rather than actual behaviour is unavoidable since we do not directly measure actual reporting behaviour in our survey study, nor can we ask about past behaviour since motivation activation is induced during the survey. However, whilst behavioural intentions are a relatively strong predictor of behaviour, we acknowledge its explanatory power does have limitations (Hassan and Wright, 2020). Still, it is frequently suggested ethical behavioural intentions can teach us about actual ethical behaviour to quite some extent (e.g., Armitage and Conner, 2001). We address this further in the discussion. All items and their Dutch translations used in the survey are included in Appendix 1.

6.3.3. Survey flow

Table 2 presents eight experimental groups. We used block order randomization and question randomization within Qualtrics to randomly distribute all respondents into one group. We required the randomizer to evenly present all elements so that the experimental groups would have similar group sizes. Four treatment groups were defined, where the activation of PSM or PM was presented first, and one of the dependent variables second. Additionally, four control groups were defined, where the order of independent and dependent variables was reversed.

Table 2 *Experimental groups*

Order of presentation ► **Experimental group ▼**	**First**	**Second**
1	PSM	Colleagues[a]
2 (control)	Colleagues	PSM
3	PSM	Patients[b]
4 (control)	Patients	PSM
5	PM	Colleagues
6 (control)	Colleagues	PM
7	PM	Patients
8 (control)	Patients	PM

Note. [a]Intentions to report wrongdoings from colleagues. [b]Intentions to report wrongdoings from patients.

6.3.4. Analysis

To assess our hypotheses, we performed OLS regressions for hypotheses 1, 2, 4 and 5. Besides, we performed moderated regression analysis to assess hypotheses 3 and 6 (Hayes, 2012).

6.3.5. Preregistration

Before the analysis, this study was preregistered to clearly segregate our hypothesis-generating and hypothesis-testing phase (See List of preregistrations). This link enables access to the preregistration and additional information.

The final study deviates slightly from the preregistration. First, we decided to test the main effect of motivation on intentions to report wrongdoings separately without hypothesis, and only define hypotheses about the effect of activation. Second, for the difference between colleagues and patients only a preliminary hypothesis had been defined which we decided to change to a non-directional hypothesis. Importantly, the design of the study, all included variables and the main expectations on the effect of activation that guided the study have remained the same and were not changed after preregistration.

6.4. Results

6.4.1. Descriptive statistics

First, we tested scale reliability for PSM and PM. Cronbach's alpha scores show that in both cases, satisfactory internal consistency is reached ($\alpha > .7$). Next, we tested unidimensionality by performing a principal axis factor analysis on the four items of both PSM and PM. Sampling adequacy is verified by the Kaiser-Meyer-Olkin (KMO) measure (for PSM .82; for PM .81) and KMO values for individual items were above .78. Eigenvalues were obtained for each factor and the data show for both PSM and PM one factor has an eigenvalue larger than 1 (Table 3).

Table 3 *Scale reliability and unidimensionality*

Item	PSM Factor loadings	PM Factor loadings
1	.85	.81
2	.87	.86
3	.80	.83
4	.88	.81
Eigenvalues	2.9	2.75
% of variance	72.37	68.73
Cronbach's α	.87	.85
n	5874	5881

Second, Table 4 presents the group sizes for the experimental groups. We find proof of the success of randomization in the fact that these experimental groups do not significantly differ on gender (one-way ANOVA: $F(7, 11{,}699) = .46$, $p = .86$), age (one-way ANOVA: $F(7, 11{,}720) = .55$, $p = .8$), size of employment contract (a dummy indicating whether a person worked less than 29 hours a week or 29 hours and more) (one-way ANOVA: $F(7, 11{,}720) = .69$, $p = .68$), and type of contract (a dummy indicating whether the employment contract was temporary or permanent) (one-way ANOVA: $F(7, 11{,}720) = 1.44$, $p = .19$).

Table 4 *Group sizes*

Type of motivation	**PSM**		**PM**		**Total**
Dependent variable	Intentions to report wrongdoings from colleagues	Intentions to report wrongdoings from patients	Intentions to report wrongdoings from colleagues	Intentions to report wrongdoings from patients	
Activated	1,506	1,397	1,490	1,500	5,893
Not activated	1,461	1,500	1,410	1,464	5,835
Total	2,967	2,897	2,900	2,964	11,728

Third, Table 5 presents the means and standard deviations for all included variables per experimental group. Further proof for the success of randomization is provided because there is no significant difference between the PSM mean of group 1 versus group 2 ($F(1, 2{,}965) = .93$, $p = .33$), and group 3 versus group 4 ($F(1, 2{,}895) = .27$, $p = .60$), and there is no significant difference between the PM mean of group 7 versus group 8 ($F(1, 2{,}962) = .64$, $p = .42$). However, there is a significant difference between group 5 versus group 6 ($F(1, 2{,}898) = 6.14$, $p = .01$). Nevertheless, this does not problematize our argument as there is no significant difference between these groups on the dependent variable. Finally, there is no significant difference between the 'Colleagues mean' of groups 1 and 2 versus groups 5 and 6 ($F(1, 5{,}865) = .19$, $p = .67$), and there is no significant difference between the 'Patients mean' between groups 3 and 4 versus groups 7 and 8 ($F(1, 5{,}859) = 1.01$, $p = .32$).

Table 5 *Means and standard deviations for all included variables*

Variable ► **Experimental group ▼**	**PSM** ***M (SD)***	**PM** ***M (SD)***	**Intentions to report wrongdoings from colleagues** ***M (SD)***	**Intentions to report wrongdoings from patients** ***M (SD)***
1	4.07 (.59)		3.89 (.97)	
2 (control)	4.05 (.63)		3.84 (1.05)	
3	4.05 (.60)			4.14 (.87)
4 (control)	4.04 (.60)			4.02 (.98)
5		4.20 (.59)	3.89 (1.00)	
6 (control)		4.15 (.61)	3.86 (1.03)	
7		4.20 (.60)		4.16 (.91)
8 (control)		4.18 (.60)		4.04 (.97)

Finally, we tested the underlying assumption of the study that motivation affects intentions to report wrongdoings in the first place, and we find PSM and PM both are significantly correlated with the intentions to report wrongdoings from both colleagues and patients (Table 6).

Table 6 *Correlation of motivation with intentions to report wrongdoings*

	Independent variable	**Dependent variable**	***B***	**sig.**	***r***
1	PSM	Intentions to report wrongdoings from colleagues	.25	.00	.15
2	PSM	Intentions to report wrongdoings from patients	.27	.00	.17
3	PM	Intentions to report wrongdoings from colleagues	.29	.00	.17
4	PM	Intentions to report wrongdoings from colleagues	.30	.00	.19

6.4.2. Main results

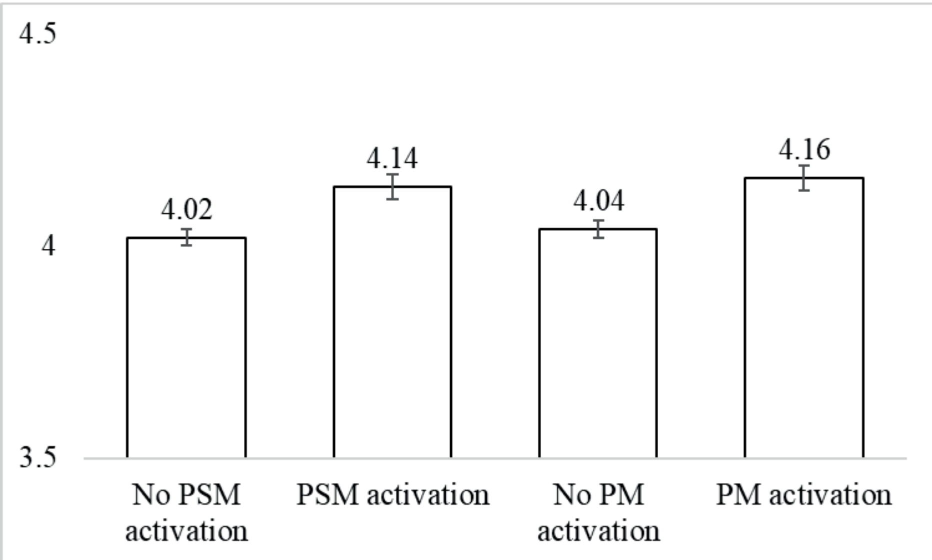
The results of the OLS regressions assessing hypotheses 1, 2, 4 and 5 are presented in Table 7. They show activating PSM as well as PM does increase intentions to report wrongdoings from patients but activating PSM as well as PM does not increase intentions to report wrongdoings from colleagues. Models 2, 4, 6 and 8 include age and gender as control variables. Models 5 and 7 in Table 7 (also depicted in Figure 1) show that intentions to report wrongdoings from patients increase with .11 or .12 on a 5-point scale when PSM or PM are activated.

Table 7 *Coefficients and standard errors of OLS regressions*

Dependent variable	Intentions to report wrongdoings from colleagues				Intentions to report wrongdoings from patients			
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Intercept	3.84* (.03)	3.75* (.11)	3.86* (.03)	3.87* (.11)	4.02* (.02)	4.07* (.09)	4.04* (.02)	3.88* (.10)
Activation of PSM	.05 (.04)	.05 (.04)			.11* (.03)	.12* (.03)		
Activation of PM			.03 (.04)	.03 (.04)			.12* (.03)	.12* (.04)
Age		.002 (.002)		-.00 (.002)		-.002 (.002)		.001 (.002)
Gender (female)		-.003 (.05)		-.006 (.05)		.04 (.05)		.12* (.05)
n	2967	2967	2900	2900	2897	2897	2964	2964
R^2	.001	.001	.000	.000	.004	.005	.004	.006

Note. * $p < .01$

Figure 1 *Graphical display of values of intentions to report wrongdoings from patients*



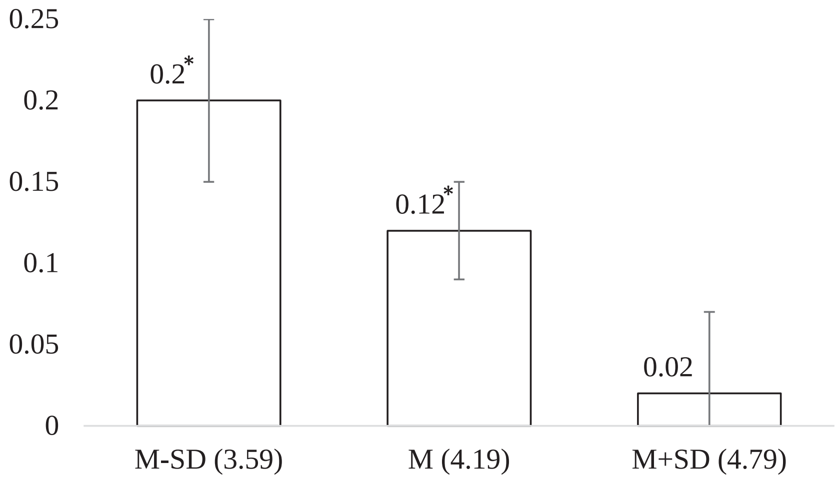
Next, we assessed hypotheses 3 and 6 with moderated regression analysis. We tested the conditional effects of PSM and PM activation on intentions to report wrongdoings from patients when using the level of PSM or PM as a moderator. We did not test the effects on intentions to report wrongdoings from colleagues as the main effect was not significant. Table 8 shows the results. Interestingly, both interaction coefficients are negative. Whereas for PSM, the interaction coefficient is not significant, the coefficient for PM shows a significant, negative interaction. This, contrary to our expectations, points out activating PM has a larger effect on intentions to report wrongdoings from patients for respondents with lower levels of PM. PSM shows the same, yet non-significant, tendency.

Table 8 *Moderation analysis for intentions to report wrongdoings from patients. Coefficients and standard errors OLS regressions*

Model	**1** **β (*SE*)**	**2** **β (*SE*)**
Intercept	2.77* (.16)	2.47* (.17)
Activation of PSM	.48** (.23)	
Activation of PM		.74* (.24)
PSM	.31* (.04)	
PM		.38* (.04)
Activation of PSM*PSM	-.09 (.06)	
Activation of PM*PM		-.15* (.06)
n	2897	2964
R^2	.03	.04

Within the moderated analysis, we conducted two further tests for the effect of PM on intentions to report wrongdoings from patients. First, we tested the conditional effect of PM activation at three values of PM (at the mean, one standard deviation under the mean, and one standard deviation above the mean). Figure 2 shows the results. The effect is significant at the two lower categories of levels of PM, but not at the highest category (*M + SD).*

Figure 2 *Conditional effects of PM activation. * p < .01*



Finally, we defined the Johnson-Neyman significance region (Hayes, 2012), which presents the value of the PM at which the effect of PM activation on intentions to report wrongdoings from patients turns (in)significant. We find the effect of PM turns insignificant (on default the significance level is set at $p = .05$) above 4.47 where 5 is the highest possible value (Mean: 4.19, $SD = .60$). 35.8% of the observations in the dataset are positioned above this value, and 64.2% below.

6

6.5. Discussion

This study has investigated whether the intentions to report wrongdoings is affected by other-regarding types of motivation (i.e., PSM and PM) and whether the question who the wrongdoer is matters in this context. Hence, the goal of this study is twofold. First, we aim to contribute to research studying ethics by paying attention to (a) motivation as a vehicle to approach intentions to report wrongdoings and b) the wrongdoer: colleagues versus patients. Second, we aim to contribute to the motivational literature by testing the effect of two other-regarding concepts, PSM and PM, on the same outcome variable—i.e., the intentions to report unethical behaviour—using the innovative measure of randomizing question order, which allows us to draw causal conclusions.

We find that the intentions of healthcare employees to report wrongdoings from patients is positively affected by the activation of both PSM and PM, providing support for ***H1*** and ***H4-A*** (and not for ***H4-B***, which states that PM decreases the intentions to report wrongdoings from patients). However, effects are small. Second, we find that the effect of PSM and PM activation is not significant when the wrongdoings are caused by colleagues. This means ***H2***, ***H5-A*** and ***H5-B*** are not corroborated. Third and final, highly interesting but contrary to our expectations, activating PM has a significantly larger effect for respondents with lower levels of PM. PSM has a non-significant larger effect for respondents with lower levels of PSM. Hence, both ***H3*** and ***H6*** are not corroborated. We discuss two main contributions of this study and their implications for practice and provide explanations for the unexpected findings. Before drawing some final conclusions, we address the limitations of this study and provide avenues for future research.

First, we contribute to the stream of research on ethical behaviour, intentions, and motivation as a driving force (e.g., Olsen et al., 2019; Ripoll, 2019; Ripoll and Schott, 2020; Steen and Rutgers, 2011), by showing that the relationship between these concepts is not universal but rather bound to context. Specifically, our results suggest this relationship is more likely to be found in an employee-patient relationship than in an employee-employee (peer) relationship. This finding is in line with the related literature on whistleblowing emphasizing the importance of the reporter-wrongdoer relationship in the intentions to report (e.g., King and Hermodson, 2000; De Graaf, 2010; Zipparo, 1999). In addition, our results can be explained by insights from the professionalism literature. In particular, the neo-Weberian approach towards professionalism warns us that professionals—like healthcare workers in our sample—can be collectively self-interested, trying to maintain a monopoly on providing certain services (Andersen and Petersen, 2012).The frequently studied phenomenon of ingroup favouritism (e.g., Ashforth and Meal, 1989; Hogg, 2018) suggests that people view their 'in-group' as deserving its successes and not its failures, while the opposite obtains for the 'out-group'. Togethers these streams of literature suggest that the bond with one's peers may override the desire to help society and (groups of) others, thereby affecting the relationship between ethics and motivation. This conclusion is indirectly supported by the findings of Stazyk and Davis' study (2015) on the relationship between PSM and ethical decision-making. The results of their large-*n* study show that only for less professionalized employees PSM appears to be positively related with ethical obligations rooted in virtue and obligation.

Second, this study contributes to the motivational literature, by nuancing Meyer-Sahling's et al. (2019) basic assumption that it is possible to activate motivation. While Meyer-Sahling's study was based on a sample of Chilean central government employees, we used data collected among Dutch healthcare employees. Given the contextual and institutional sensitivity of PSM (Vandenabeele and Van der Wal, 2008), this study shows that previous conclusions on the effectiveness of activation may be generalized but that the sizes of the effects are small. In addition, our results show that the possibility to activate motivation is not unique to PSM but also applies to PM. We thereby contribute to the small yet growing number of studies that study the interrelatedness and outcomes of PSM and PM, (e.g., Peng and Li, 2019; Piatak and Holt, 2019; Ritz et al., 2020; Wright and Christensen, 2013). In particular, the similar effect sizes of PSM and PM activation reinforce past findings suggesting that there may be quite some empirical overlap between PSM and PM when using the global PSM measurement. However, scholars using qualitative data (Schott et al., 2019) or quantitatively measuring PSM-dimensions separately (Ritz et al., 2020) do find distinctions between PSM and PM. This shows that the question of how PSM should be measured remains one of the big challenges of PSM research (Perry and Vandenabeele, 2015; Vandenabeele and Schott, 2020).

What is more, contrary to our expectations based on Meyer-Sahling, Mikkelsen, and Schuster (2019), but in line with Linos (2018), we find PM activation has a larger effect for respondents with lower rather than higher levels of PM (for PSM, we find a similar but smaller and non-significant effect). There may be valid reasons why our results diverge this drastically. Meyer-Sahling, Mikkelsen, and Schuster (2019) obtained lower means of PSM in their sample consisting of Chilean civil servants (3.55 for PSM) than we do in our sample of Dutch healthcare employees (4.06 and 4.05 for PSM, and 4.18 and 4.19 for PM in our subsamples). Because we retrieve higher means this may enable us to observe 'ceiling effects' of activation, whereas Meyer-Sahling could not. A ceiling effect would suggest that at certain high levels of motivation, activation is ineffective because there is little room for improvement. This argument is reversable. It is possible Meyer-Sahling, Mikkelsen, and Schuster (2019) were able to observe 'floor effects': at certain low levels of PSM, activation is ineffective because there is no PSM to activate.

Although effects are small in our survey setting, the preliminary confirmation that motivation can be activated should open avenues for (behavioural public administration) scholars to consider whether activation maybe more effective in other settings. We suggest considering motivation activation as a potentially effective behavioural intervention. Such an intervention would foster behavioural change through the activation, or reminding of, a related concept. This insight shows resemblance to literature on choice architecture as well as on micro interventions. In a taxonomy on choice architecture techniques, Münscher, Vetter, and Scheuerle (2016) present 'providing reminders' as one of the possible techniques. They argue 'choice architects can intervene by providing positive reminders that heighten the salience of a desired option' (519). However, where they limit the options to 1) reminding someone of the concept that is to be stimulated (e.g., voting), or 2) oppressing reminders of concepts that are to be discouraged (e.g., eating unhealthy food), our study presents a new subtype of reminders where the concept that we remind someone of (PSM or PM) is different from, but related to, the behaviour that we aim to change (ethical reporting). A recent study used a similar approach when discussing micro interventions. The specific intervention made participants recall their prosocial or societal impact, and measured effects on, among other, turnover intention (Vogel and Willems, 2020).

6.5.1. Practical implications

Motivation activation is an interesting potential management tool to stimulate ethical considerations and behaviours among employees. For example, in formal and informal communication to their employees about the importance of ethical reporting, managers can remind employees of their motivation to do good or help others, which is likely to increase (intentions of) ethical reporting. Herein, the literature on choice architecture in general and providing reminders specifically can assist regarding ideas on how to design such reminders. However, our study also reveals the limits of motivational activation in the context of professional work. Behavioural interventions directed at peer reporting through activating motivation may be less effective than interventions directing at patient reporting. These peer-directed interventions could benefit from other behavioural interventions, e.g., through framing ethical reporting as an expression of collective self-interest ('reporting a peer is in the best interest of all peers') rather than peer betrayal. Finally, our observed 'ceiling effects' of activation suggests that behavioural interventions are most effective when directed at employees with medium motivation scores: those who already know the way but need a gentle push.

6.5.2. Limitations

Notwithstanding the above contributions, this study has some limitations. First, we should be somewhat careful comparing our study to Meyer-Sahling, Mikkelsen, and Schuster (2019) since we use slightly different operationalizations of the key variables. In particular, we use a shorter unidimensional PSM scale and present two adopted outcome variables. Using the shorter, unidimensional PSM scale as advantages and disadvantages. It presents as good as a measure of PSM as a multidimensional alternative (Wright, Christensen, and Pandey, 2013), and may be more able to grasp the core driver of the concept (Vandenabeele, Ritz, and Neumann, 2018). A shortcoming of the unidimensional scale, in contrast, may be its reduced strength as a prime, which may partly explain the small effect size found of in this study.

Second, similarly to Meyer-Sahling, Mikkelsen, and Schuster (2019), we do not measure actual behaviour but intent. We acknowledge measuring intentions does not allow us to draw strong conclusions concerning real live behaviour because intentions do not equal behaviour. On the one hand, studies do find intentions do reflect behaviour at least to some extent (Armitage and Conner, 2001; Hassan and Wright, 2020). On the other hand, especially for behaviour related to whistleblowing, intentions and behaviour may diverge (Mesmer-Magnus and Viswesvaran, 2005). A real whistleblowing situation introduces many variables not present in a simulation, like experiencing emotions or external pressure (Fishbein and Ajzen, 2010). Therefore, it is important to be clear about what we can and cannot learn from this study. We cannot prove that the patterns we found are equally present in actual behaviour, and it is important future studies investigate this link. However, our study does introduce and measure the effect of new factors, PSM versus PM and colleagues versus patients, on a large sample, which increases our understanding of intentions to report wrongdoings and opens new avenues to study ethical behaviour (Bjørkelo and Bye, 2014). Related to this, we should acknowledge social desirability bias (SDB) may played a role in our study, which is argued to be a risk in PSM studies (Kim and Kim, 2016). However, we think there are strong grounds to believe that the treatment manipulation was successful and not just a result of question order. If effects were based on question order, then it would be likely to observe effects of this in the control groups, where the order of independent and dependent variables was reversed. In other words, in the control groups, question order effects should then be observed in the independent variable (PSM or PM), as respondents may be affected by social desirability bias due to the exposure to the dependent variable, the question on intentions to report wrongdoings. However, we do not see large differences in the means of PSM or PM across control groups and treatment groups, supporting our argument that the treatment was successful.

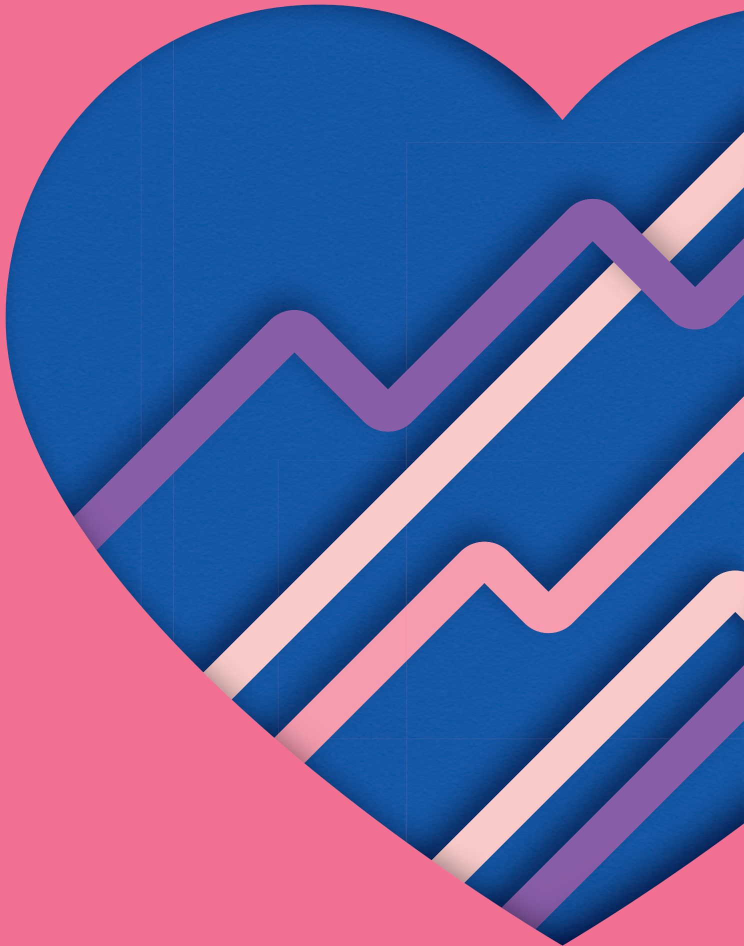
Third, the observed relationships have low explained variance meaning that our activation effects are small. This observation finds resemblance in the extant literature. First, we find similar observed effects compared to Meyer-Sahling, Mikkelsen, and Schuster (2019). At the same time, we need to acknowledge other studies who did not find a relationship between PSM priming and unethical behaviour (Olsen et al., 2019; Christensen and Wright, 2018), suggesting that more research is needed before strong conclusions can be drawn. Second, arguably, activation of PSM or PM is not nearly the sole explanator of intentions to report wrongdoings. In our study, however, we have not aimed to investigate into the many or most important factors that may influence intentions. Rather, we wanted to explore subtle differences regarding the activation of types of other-regarding motivation (PM or PSM) and who is being reported.

6.5.3. Future research

Our study opens several avenues for further research. First, interesting questions arise surrounding the boundary conditions that affect the effectiveness of motivational activation. In this study we have differentiated between reporting on peers and reporting on patients/clients within the healthcare sector. As pointed out above, the strong professional identity of individuals working in this sector may explain our finding. Next to medicine, engineering and law present examples of classic professions with a strong identity (Krause, 1996). We therefore think that it is reasonable to assume that our results can be generalized to public service arenas where public servants with a strong professional identity work. In future research projects, we encourage scholars to test this assumption. Also note that we left it to respondents themselves to interpret what groups were referred to. Therefore, we should acknowledge that across organizations and even across employees, interpretations of what a patient or client is, and what nature their dependency, may differ. An interesting avenue would be finding out whether different interpretations of the same concept matter for the effectiveness of activation. Looking beyond peer versus patient, another important wrongdoer can be the employing organization (De Graaf, 2010), with management as the responsible party. Given the extra-organizational focus of PSM (Vandenabeele, 2007), we encourage scholars to investigate whether motivational activation works for this type of wrongdoer, too. A related promising avenue for future research may be to investigate what the relationship between motivational activation and the intentions to report wrongdoings looks like for less professionalized employees. Stazyk and Davis' large-*n* study (2015) on the relationship between PSM and ethical decision-making showed that only for less professionalized employees PSM appears to be positively related to ethical obligations rooted in virtue and obligation. The degree of professionalization may help to explain why we find that motivational activation only effects the intentions to report wrongdoings from patients, but not those caused by colleagues.

Second, we have only started exploring the effects of different types of motivations on intentions to report wrongdoings, and with mixed success. This begs the question of what concepts can and cannot be activated and what outcomes can and cannot be affected by what concepts. For example, is it possible to influence actual ethical behaviour (rather than intentions) by activating motivational concepts? How do the different motivational concepts relate (e.g., do PSM and PM correlate and how does this affect their distinct effects)? And can we influence behaviours other than ethical behaviours relevant in the public sphere, such as co-producing (Meijer, 2011) or prosocial behaviour (Awan, Esteve and Witteloostuijn, 2020)? To answer the above questions, we need to employ other methods. For example, qualitative research can be used to generate rich data and further explore the theory of ethical behaviour and motivation and its underlying mechanisms. Also, we need experimental research to causally assess the relationships in lab or field settings, as this allows to measure actual behaviour. Specifically, field experiments could develop real-life interventions that do not only further scientific knowledge on the topic but also present valid choice architecture techniques for practitioners (Hansen and Tummers, 2020).

Concluding, public servants are expected to maintain high ethical standards, but in doing so their management can deliver necessary support. To prevent employees from allowing unethical behaviour, activation of motivation is a potential avenue to stimulate the intentions to report ethical issues. However, the effectiveness of this approach turns out to be dependent on who the wrongdoer is and what level and type of motivation is present within the reporter. Additionally, the relatively small impact that activation makes in existing studies begs for different approaches to study motivational activation. All in all, this study presents fruitful avenues for researchers and practitioners alike.



Chapter 7

Autonomy-preserving nudges to reduce email use

This chapter is based on the following published article:
Van Roekel, H., Giurge, L., Schott, C., & Tummers, L. (2023). Nudges can be both autonomy-preserving and effective: Evidence from a survey and quasi-field experiment. *Behavioural Public Policy*. DOI: 10.1017/bpp.2023.18

Abstract

Nudges are widely employed tools within organizations, but they are often criticized for harming autonomy and for being ineffective. We assess these two criticisms simultaneously: can nudges be both autonomy-preserving and effective in changing behavior? We developed three nudges—an opinion leader nudge, a rule-of-thumb and self-nudges—to reduce a particularly sticky behavior: email use. In a survey experiment of 4,112 healthcare employees, we tested their effect on perceived autonomy and subjective effectiveness. We also tested traditional policy instruments for comparison. Next, to assess objective effectiveness, we conducted a quasi-field experiment in a large healthcare organization with an estimate of 1,189 active email users. We found that each nudge in isolation, but especially when combined, was perceived to be both autonomy-preserving and effective, and more so than traditional policy instruments like an access limit or a monetary reward. We also found some evidence that the combination of all nudges decreased actual email use. This paper advances the literature by showing how innovations in nudge design improve nudges' ability to be autonomy-preserving and effective.

7.1. Introduction

Nudges have become widely employed tools within organizations. The popularity of nudges is largely attributed to their advantages: they are simple and low-cost to implement. A nudge is an intervention that alters people's behavior by changing the choice architecture in which a decision is made. Importantly, it does not forbid anything or change incentives (Thaler and Sunstein, 2021). Nudges work by altering information, changing the structure of a decision or assisting with decision-making (Münscher et al., 2016). They have been shown to increase healthy life choices (Lin et al., 2017), stimulate evidence-based medicine (Nagtegaal et al., 2019) and improve human-computer interaction (Caraban et al., 2019).

However, nudges have also attracted criticism. Two critiques are particularly salient in the scholarly literature and in the media (Tummers, 2022). First, scholars have argued that nudges reduce autonomy (Hausman and Welch, 2010; Wilkinson, 2013). Although some studies have addressed this criticism (e.g., Wachner et al., 2020, 2021), the debate continues in part because scholars use varying conceptualizations of autonomy (Vugts et al., 2020). In this paper, autonomy is understood as the extent to which employees experience agency: to what extent does a nudge allow employees to make independent decisions in their work? This approach fits the findings of Vugts et al. (2020), who show most scholars, albeit implicitly, understand nudge autonomy in terms of agency. Autonomy is a fundamental human need that drives motivation (Ryan and Deci, 2017). If nudges decrease autonomy, their presence may not be desirable. The second criticism is that nudges are ineffective in changing behavior. In defense of nudges, a recent meta-analysis conducted on more than 200 studies found that nudges are effective with small to medium effect sizes (Mertens et al., 2022). However, other scholars find no evidence for the effectiveness of nudges after correcting for publication bias (Maier et al., 2022). Relatedly, as Mertens et al. (2022) admit, a nudge's effectiveness often depends on the type of nudge. Finally, there is scant knowledge about why nudges are effective (Szaszi et al., 2022).

These two criticisms should urge scholars to study whether nudges are autonomy-preserving and effective as well as why, how and under what conditions they work. Autonomy and effectiveness of nudges may present a tension: effective nudges could be less autonomy-preserving and vice versa. For example, defaults are more effective than other types of nudges (Mertens et al., 2022), but respondents also expected default nudges to be particularly detrimental to autonomy (Wachner et al., 2020, 2021). However, this tension between autonomy and effectiveness is not a given. An effective nudge can also increase autonomy by helping people make the choices they want to make (De Ridder et al., 2020). We argue that whether nudges can preserve autonomy and be effective at the same time depends on the nudge design. Recent innovations in nudge theory, like those on nudge+, nudge vs think, boosting and self-nudges, can inform the development of nudges that are autonomy-preserving and effective (Hertwig and Grüne-Yanoff, 2017; Reijula and Hertwig, 2022; Banerjee and John, 2023).

Building on nudge theory innovations, we developed three nudges—an opinion leader nudge, a rule-of-thumb and multiple self-nudges—that target a sticky behavior: email use. Prior research has shown that despite its promised benefits, email has become a source and symbol of stress at work (Barley et al., 2011; Brown et al., 2014). As a result, email has been associated with a host of negative outcomes, including lower work quality (Rosen et al., 2019), increased burnout threat (Belkin et al., 2020) and decreased life satisfaction (Kushlev and Dunn, 2015). Reducing email use has therefore become a topic of increasing attention among scholars and practitioners (Cecchinato et al., 2014; Bozeman and Youtie, 2020).

To what extent can nudges preserve autonomy and be effective in decreasing email use? First, in a pilot study of 435 employees, we tested whether the three nudges are perceived as autonomy-preserving and effective. Next, in a large-scale survey experiment among 4,112 healthcare employees, we measured perceived autonomy and subjective nudge effectiveness in comparison to traditional email interventions based on policy instruments like a monetary reward. Because social desirability bias can threaten the validity of a survey, we added a modified version of the Bayesian truth serum to illicit more truthful responses (Prelec, 2004). Finally, to test for objective nudge effectiveness, we implemented a quasi-field experiment in a large healthcare organization with an estimate of 1,189 active email users.

Overall, by showing that we can design nudges that are perceived as autonomy-preserving and effective, our paper provides much-needed nuance to the debate surrounding nudge development (Wilkinson, 2013; Wachner et al., 2021; Mertens et al., 2022). We also contribute by showing how nudges could help reduce email use. Email communication has long posed a threat to employee productivity and well-being and existing research has failed to provide a solution (for an exception, see Giurge and Bohns, 2021). Finally, our paper makes a concrete methodological contribution by showing how nudges can be tested using both perceptions and behavioral outcomes. We also include multiple insights that enrich the results, by, for example, using a Bayesian truth serum to counter social desirability bias and comparing nudges to traditional policy instruments (Prelec, 2004; Tummers, 2019).

7.2. Theory

7.2.1. The nudge debate

In their influential book *Nudge* (2008), Thaler and Sunstein describe how organizations can use nudges to cope with biases in human decision-making. Rooted in behavioral economics, a nudge is an intervention that aims to influence people's behavior based on insights about the bounded rationality of people (Hansen, 2016). Bounded rationality refers to the notion that people are imperfect decision-makers that do not have access to all information and computational capacities that are required to evaluate the costs and benefits of potential actions. It contrasts the rational agent model prevalent in neoclassical economics, in which people are seen as rational agents that maximize their utility (Simon, 1955). Building on the work of Simon, psychologists aimed to develop maps of bounded rationality (Tversky and Kahneman, 1974, 1981; Kahneman, 2011). They analyzed the systematic errors that distinguish the actions people take from the optimal actions assumed in the rational agent model. Tversky and Kahneman (1974) show that heuristics, though useful, can lead to predictable and systematic errors. For example, anchoring bias refers to the tendency of people to overvalue the first piece of information they receive (e.g., Nagtegaal et al., 2020). Such biases explain why people sometimes do not respond to traditional managerial instruments, like a bonus or a ban (Tummers, 2019). Instead, nudges aim to change the choice architecture without changing economic incentives or forbidding any options (Münscher et al., 2016; Thaler and Sunstein, 2021).

Despite their popularity, nudges are not without criticism, two of which are particularly salient (see e.g., Tummers, 2022). First, nudges are said to reduce autonomy. Indeed, some scholars argue that because nudges work via unconscious processes, they can exploit weaknesses, manipulate and reduce choice (Hausman and Welch, 2010; Hansen and Jespersen, 2013; Wilkinson, 2013), and as a result harm autonomy. Nudges that are not autonomy-preserving are problematic because autonomy presents one of three basic and universal psychological human needs (next to the need for relatedness and the need for competence) that drives human behavior and motivation (Ryan and Deci, 2017). Notably, there are also scholars who argue choice architecture is always present, regardless of whether one actively influences it (Sunstein, 2016). Empirical evidence on this issue is inconclusive. Studies suggest some nudges can harm autonomy while others do not (e.g., Wachner et al., 2020, 2021, Michaelsen et al., 2021). Similarly, research on the public acceptance of nudges also shows mixed findings (Davidai and Shafir, 2020; Hagman et al., 2022).

Besides the criticism on autonomy, scholars have disputed whether nudges are effective. In a recent meta-analysis, with over 200 studies, Mertens et al. (2022) found that nudges are, on average, effective in changing behaviors with small to medium effect sizes. There are, however, several counterarguments to this claim. First, the authors indicate that the effectiveness of a nudge depends on the type of nudge: nudges focused on decision structure (e.g., default nudges), outperform nudges focused on decision information or decision assistance. Second, Szaszi et al. (2022) note that context matters: whether nudges are effective varies and the conditions under which they work are barely identified. Third, Maier et al. (2022), in a response to Mertens et al. (2022), point out that after correcting for publication bias, there is no evidence that nudges are effective. Related to this this debate, Bryan et al. (2021) note that instead of focusing on replication in behavioral science, we need a heterogeneity revolution by analyzing which particular nudge works for what situation.

7.2.2 Autonomy-preserving and effective nudges

One of the reasons for the different views on autonomy and nudging depends on one's definition of autonomy, or lack thereof. Based on a systematic review, Vugts et al. (2020) show that the discussion surrounding nudge autonomy is clouded by different conceptualizations of autonomy. They identify three conceptualizations of autonomy (p. 108), namely freedom of choice (i.e., 'the availability of options and the environment in which individuals have to make choices'), agency (i.e., 'an individual's capacity to deliberate and determine what to choose'), and self-constitution (i.e., 'someone's identity and self-chosen goals'). A nudge could simultaneously decrease one's autonomy in one conceptualization and increase one's autonomy for another (Vugts et al., 2020). For example, by limiting one's freedom of choice you could help people reach implicit goals and improve self-constitution. Similarly, a nudge can help someone think about the right choice but also limit the range of available choices to pick from—this would promote agency but limit freedom of choice.

In this paper, we adopt the conceptualization of autonomy as agency because it presents a higher threshold for nudge autonomy than the initial definitions of nudging and libertarian paternalism (see Thaler and Sunstein, 2008). The initial understanding of autonomy in the nudge literature relied heavily on freedom of choice, but 'apart from a context that allows choice, autonomy also requires a capacity to choose and decide' (Vugts et al., 2020: 116).

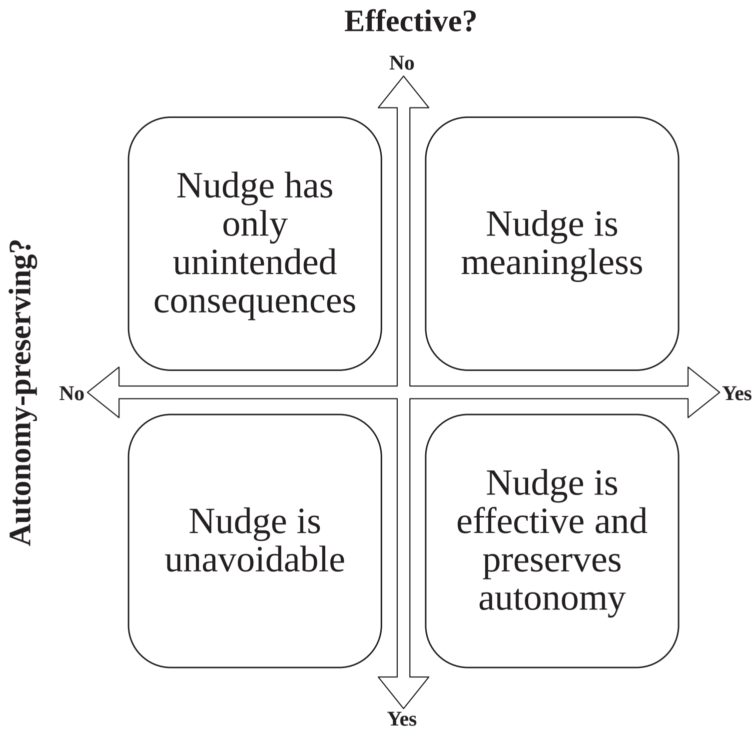
At first glance, autonomy and effectiveness may appear to present a tension. For a nudge to be autonomy-preserving, the assumption is that the nudge guarantees agency (Vugts et al., 2020)—meaning that it allows someone to execute their personal judgement (Morgeson and Humphrey, 2006; Gorgievski et al., 2016). In contrast, an effective nudge assumes someone's personal judgement is flawed, because nudges are effective by being based upon—and making use of—biases in human decision-making (Hansen, 2016). While not removing any option, a nudge is effective by actively changing the choice architecture (Thaler and Sunstein, 2021) and may be considered manipulative (Wilkinson, 2013). For example, decision-structure nudges like defaults are more effective than decision information nudges like social norms (Mertens et al., 2022). At the same time, people also expect default nudges to lower autonomy more so than social norm nudges (Wachner et al., 2020; Wachner et al., 2021). In contrast, people find social norm nudges to be more autonomy-preserving, yet these are more often ineffective (Wachner et al., 2021). Although in some cases the autonomy-effectiveness tension may emerge, it is not a given. In fact, scholars have developed arguments about how autonomy and effectiveness go hand in hand. For example, De Ridder et al. (2020) argue an effective nudge can increase autonomy by helping people make the choices they want to make.

In Figure 1, we consider four scenarios for the ability of nudges to preserve autonomy and be effective. These scenarios are theoretical extremes and do not suggest that it is a yes/no question. First, when a nudge is effective but not autonomy-preserving, it is unavoidable and may well be a nudge that appears manipulative in the sense that it does not offer 'an escape clause' (Wilkinson, 2013: 354). Second, when a nudge is neither autonomy-preserving nor effective, it can decrease autonomy, while failing to do what it intended to do. Other unintended consequences may include when a nudge backfires or triggers reactance (Osman, 2020). Third, when a nudge is not effective but preserves autonomy, the nudge is meaningless and might be met with indifference (Wachner et al., 2021). Finally, nudges could preserve autonomy and be effective (De Ridder et al., 2020).

It is likely that the specific design of nudges impacts their autonomy and effectiveness. As such, when designing nudges, scholars should make use of theoretical concepts that are expected to increase nudges' ability to preserve autonomy and be effective. Below, we discuss nudge vs think, boosting and self-nudges (Sunstein, 2016; Hertwig and Grüne-Yanoff, 2017; Reijula and Hertwig, 2022).

First, scholars have discussed how nudging relates to thinking. Sunstein (2016) distinguishes System 1 (based on fast, intuitive thinking) and System 2 (using slow deliberation) nudges (based on Kahneman, 2011). People seem to prefer System 2 nudges (Sunstein, 2016). Furthermore, Lin et al. (2017) argue that nudges that promote reevaluation are more effective. Therefore, nudges that rely on conscious (System 2) decision-making, for example through employing information, may combine autonomy and effectiveness, more so than nudges that rely on unconscious (system 1) processes, for example by changing defaults. Additional arguments about how nudging can include deliberation have been developed by John et al. (2009) who discuss how 'nudge' and 'think' as behavioral change strategies may influence each other. Herein, a deliberative nudge refers to the combination of both (John et al., 2023). In a similar vein, nudge+ refers to a nudge that has a reflective element embedded in its design (Banerjee and John, 2021; 2023).

Figure 1 *Four scenarios for autonomy and effectiveness of nudges*



Second, scholars have advocated to use boosting as a behavioral intervention (Hertwig and Grüne-Yanoff, 2017; Hertwig, 2017). A boost aims to enhance someone's decision-making by providing skills, knowledge or tools. Those advocating for boosting go even further than System 2 nudges by arguing that bounded rationality is malleable, and interventions should teach the decision-maker to change their behavior. An example of a boost is improving statistical reasoning with a brief training (Bradt, 2022). Like the distinction between System 1 and System 2 nudges, in practice, the distinction between boosts and nudges may be harder to uphold and interventions may carry characteristics from both nudging and boosting (Van Roekel et al., 2022). Yet, we can leverage this conceptual overlap by developing nudges that use insights from boosting. For example, to introduce a decision tree to guide, decision-making is a type of boost (Hertwig and Grüne-Yanoff, 2017). A rule-of-thumb, which is a type of nudge, is effectively a simpler version of a decision tree (Münscher et al., 2016).

Finally, scholars have been studying involvement of people in designing choice architecture. Involving employees in the process may make nudges more autonomy-preserving and effective. When it comes to development, this means being transparent and involving the target group in the trajectory of analyzing behavior and designing the nudges to get their support (Bruns et al., 2018; Tummers, 2019). Employees could also be involved in the execution, an insight derived from the concept of self-nudging. Self-nudging suggests people can use nudges to self-regulate: 'nudger' and 'nudgee' become the same person. Self-nudges require awareness of how one's environment affects one's behavior as well as knowledge of a nudge that can modify this relationship (Reijula and Hertwig, 2022: 123). In that sense, self-nudges can be regarded as a type of boost and may present a type of behavioral intervention that is both autonomy-preserving and effective (Reijula and Hertwig, 2022).

7.2.3. The role of nudges in reducing email use

We study nudges in the context of email use. Email has become a primary means of communication at work, because, in theory, it allows employees to decide when and where to work (Rosen et al., 2019). However, email communication has become a unique job demand and source of stress at work because it facilitates non-stop sharing or requesting of input (Barley et al., 2011). As such, many employees today feel compelled to read and respond to email in real-time, contributing to the development of workplace norms around continuous connectivity and instant responsiveness (Brown et al., 2014; Giurge and Bohns, 2021).

Because of these norms, email communication has been associated with a host of negative consequences both at work and outside of work. For example, email undermines work quality because it fragments employees' attention (Jackson et al., 2003). Email communication has also been associated with greater burnout and lower life satisfaction in part because it prohibits employees to disconnect from work and engage in non-work activities such as leisure that are beneficial for well-being (Kushlev and Dunn, 2015; Belkin et al., 2020).

Ironically, although many studies tend to use email to distribute nudges (DellaVigna and Linos, 2022), nudges have rarely been used to directly alter email use. Rather, most studies that aim to alter email use are focused on changing the person rather than the environment. For example, Dabbish and Kraut (2006) found that several individual email management tactics (e.g., having less email folders) were associated with lower email overload. Relatedly, Gupta et al. (2011) found that limiting the moments when one checks their email decreased stress, which in turn predicted greater well-being. In terms of nudges, we only found one paper that employed nudges to increase awareness of phishing (Vitek and Syed Shah, 2019), which arguably focuses more on changing how employees interact with the content of emails than with email use. However, outside the academic literature, many nudge-like software is available, like reminders when an email is written poorly or a simple cognitive test to assess whether the user is fit to send emails at certain times (Balebako et al., 2011). In line with this evidence, we expect that nudges can be used not only to inform people about how to improve and engage with email content but also how to address email use altogether (Cecchinato et al., 2014; Bozeman and Youtie, 2020). Hence, our main hypothesis[1] is:

H1: *Nudges will be both autonomy-preserving and effective in decreasing email use.*

7.3. Methods

We study the autonomy-preservation and effectiveness of nudges in the context of email use among healthcare workers, which is a group of employees that are particularly prone to burnout and high email use (Reith, 2018; Van Roekel et al., 2021). Table 1 presents an overview of our studies. After the pre-study to develop the nudges and the pilot to test the nudges, the empirical studies (the survey experiment and the quasi-field experiment) allow to evaluate our main hypothesis. The main empirical studies underwent ethical review were preregistered (See List of preregistrations; see also Appendix 7) and present open data (more information below).

1 We preregistered four hypotheses: one or two hypotheses per study. In the general paper, we focus on one overarching hypothesis. Appendix 7 discusses the preregistration and evaluation of all original hypotheses.

Table 1 *Overview of studies*

Study	Method	*n*	Goal
Pre-study	Interviews	11 respondents	Develop nudges
Pilot	Survey experiment	435 respondents (general population)	Test nudges
1	Survey experiment	4,112 healthcare employees	Test ***H1*** (perceived autonomy and subjective nudge effectiveness)
2	Quasi field-experiment	Healthcare organization with an estimate of 1,189 active email users	Test ***H1*** (objective nudge effectiveness)

7.3.1. Pre-study: developing nudges

In a pre-study, we developed the nudges and interviewed 11 employees in the organization where we would later conduct our quasi-field experiment (5 HR advisors, 1 program manager, 1 team manager, 1 nurse, 1 occupational physician, 1 occupational health nurse and 1 IT employee). Appendix 1 contains the semi-structured interview guide. We used the interviews to develop three nudges: an opinion leader nudge, a rule-of-thumb and multiple self-nudges. An opinion leader nudge is a message that describes the behavior of a person of influence, assuming it will convince receivers to follow the social reference point (Valente and Pumpuang, 2007; Münscher et al., 2016). A rule-of-thumb is an easy-to-follow guideline that works well in most situations and decreases the effort of a decision (Münscher et al., 2016; Hertwig and Grüne-Yanoff, 2017). Self-nudges are nudges redesigned to be used by employees to nudge themselves. They help boost self-control (Hertwig and Grüne-Yanoff, 2017). Appendix 2 presents the nudges that we developed to decrease email use in detail.

7.3.2. Pilot: testing nudges

After developing the nudges, we piloted them in an online Prolific panel (n = 435). We used the panel to assess the perceived autonomy and subjective nudge effectiveness from the perspective of the general working population (DellaVigna et al., 2019). For the measurement of subjective nudge effectiveness, we asked the panel to predict the feasibility, appropriateness, meaningfulness, and effectiveness of the nudges in their organization (following the FAME-approach for evidence-based practice, Jordan et al., 2019). Appendix 3 presents all survey measures used in this paper and Appendix 4 describes the methods and results of the pilot in detail.

All nudges were assessed as autonomy-preserving. Respondents indicated that on average they ‘somewhat agree’ to ‘agree’ with the nudges being autonomy-preserving: the means are all above 5 on a 1-7 scale. The results also indicate respondents thought the nudges would be ‘somewhat effective’. Scores were highest for appropriateness (4.55-5.28 on a 7-point scale) and lowest for effectiveness (3.97-4.35). The only score just below the midpoint (<4) was for the effectiveness of the rule-of-thumb (3.97), indicating that this nudge was perceived as least effective.

7.3.3. Study 1: Survey experiment

The goal of Study 1 was to test perceived autonomy and subjective nudge effectiveness in a large-scale survey experiment among healthcare employees in the Netherlands (n = 4,112). Employees assessed the nudges individually and combined. To compare these nudges with alternative organizational interventions, we also asked employees to assess traditional policy instruments (i.e., an email access limit to limit emailing to only 2 hours per day, a monetary reward for emailing less than before or public praise for emailing less than your colleagues). Appendix 5 details the specific text of the traditional interventions.

The large-scale survey experiment was part of a longitudinal survey for which ethical approval was granted (Faculty’s Ethical Review Committee of the Faculty of Law, Economics and Governance, Utrecht University; no. 2019-004). Respondents provided informed consent, including allowing for the publication of anonymized data. The main data for this study are available at https://osf.io/6n2g4/?view_only=895be1c46d384867b52e22ff30892ba8.

Participants

We collected data between 18 May and 20 June 2022 via a Qualtrics survey. Respondents were required to use email in their job, list-wise deletion was applied.

The mean age of the respondents was 51.94 (SD = 9.65, Min. = 20, Max. = 74, 3 missing). Regarding gender, 3,506 were female (85.3%), 587 were male (14.3%) and 19 respondents indicated X or that they would rather not say (0.4%). Respondents worked in all healthcare sectors: 1,515 (36.8%) in hospitals, 1,059 (25.8%) in nursing or home care, 653 (15.9%) in mental healthcare, 620 (15.1%) in disabled care and 265 (6.4%) in other healthcare. A total of 2.116 (51.5%) respondents worked 29 or more hours a week, 1.914 (46.5%) of respondents worked 16-28 hours a week, 69 respondents (1.7%) worked 15 hours or less, and 13 respondents (0.3%) reported to have a zero-hours contract.

Procedure and measures

All survey measures mentioned below are included in Appendix 3. Respondents first passed an eligibility check (respondents had to use email at their job). We assessed two measures of email use (email volume and email time) with open questions adapted from Sumecki et al. (2011). We assessed email overload with a seven-item scale (α = 0.80) adapted from Dabbish and Kraut (2006), ranging from 'strongly disagree' (1) to 'strongly agree' (7). Next, respondents were exposed to one of the seven interventions randomly: one of the three nudges, the combination of all nudges or one of the three traditional email interventions (translated in Dutch). Appendix 6 shows that randomization was successful across gender, age, healthcare sector and amount of working hours. The instruction accompanying the intervention read: 'Imagine the organization you work for sends you the following message about using email in your organization. Please read the message carefully'.

After the conditions, respondents were asked to evaluate the subjective nudge effectiveness. We used self-admission rates for our main analysis and added an adapted version of the Bayesian truth serum to increase the credibility of the given answers (Prelec, 2004; John et al., 2012 Weaver and Prelec, 2013; Frank et al., 2017; Van de Schoot et al., 2021 Schoenegger, 2023). Appendix 3 elaborates on the serum.

Next, like the pilot study, we assessed perceived autonomy with three items (α = 0.93) on a 7-point Likert scale (Decision-Making Autonomy; WDQ; Morgeson and Humphrey, 2006; translation adapted from Gorgievski et al., 2016). We also measured work engagement with three items (α = 0.80) on a 5-point Likert scale ranging from 'Never' (1) to 'Always (daily)' (5) (Schaufeli et al., 2019). All items/questions were translated in Dutch. In the survey, items within each measure were randomized. At the end of the survey, respondents provided background characteristics. We used one-way analysis of variance for our main analysis. Significance levels were set at $p = 0.05$. We report exact p-levels.

7

7.3.4. Study 2: Quasi-field experiment

In Study 2, we tested the nudges sequentially in a quasi-field experiment in a large Dutch healthcare organization. This quasi-experiment had a One-Group Pre-test-Post-test Design with multiple sequential treatments and post-tests (Shadish et al., 2002).

The prefix 'quasi' is appropriate because the experiment did not include a control group (Shadish et al., 2002). Within the organization, there were technical limitations so interventions could not be randomly distributed to a selection of employees (i.e., there was no option to randomly send text messages to employees or randomly show intranet messages to a selection of employees) nor could any treatment group be separated from a control group when measuring email use. Moreover, any alternative treatment distributions that were considered (e.g., physical posters) would risk spill-over effects within the organization. The main disadvantage of a design without a control group is that differences in email use between pre- and post-intervention periods may be caused by elements or events unrelated to the treatment. Nevertheless, there are two main reasons why quasi-experiments are valuable designs to assess causality in instances where randomized designs are not possible or desirable (Shadish et al., 2002; Grant and Wall, 2009). First, a quasi-experiment like this one should be seen as a method of action research, an opportunity for collaboration between researchers and practitioners to jointly improve, in this case, the use of email within the organization (Grant and Wall, 2009). Second, our study still provides an estimate of the nudge effectiveness and we have taken several measures to improve reliability, including only measuring full workdays (excluding weekends and single holidays), planning the experiment in a period where few natural fluctuations were expected (there were no long holidays right before, during or after the experiment and no major events happened) and tracking email use for 8 weeks. Such measures make the effects of external events (like history or maturation) less likely (Shadish et al., 2002).

We tested a total of four interventions (the three nudges and their combination), added a post-test after each intervention, an additional post-test before the combination of nudges, and two additional post-tests at the end of the experiment. Email use was measured weekly for eight consecutive weeks. Consequently, our design was the following, whereby O_n refers to the *n*th test and X_n to the *n*th treatment:

$$X_1\ O_2\ X_2\ O_3\ X_3\ O_4\ O_5\ X_4\ O_6\ O_7\ O_8$$

In this study, we measured email use only, and not perceived autonomy, for two reasons. First, the added value of the field experiment was to test the effectiveness of the nudges on real behavior, while the survey experiment establishes perceptions of effectiveness and autonomy. A measure of autonomy in the field experiment would, again, be a perception. Second, the organization in which the quasi-field experiment was conducted, did not allow for large-scale surveying of employees, making it impossible to collect employees' perceptions about autonomy. A drawback of this approach is that autonomy may be perceived differently in the field compared to the survey study. While we cannot rule this out, it would only be a problem if nudges are considered less autonomy-preserving in real-life settings compared to hypothetical settings. However, a recent study showed that when people expect a nudge to diminish their autonomy in a hypothetical setting, they do not report any differences in autonomy for that same nudge in a real-life setting (Wachner et al., 2021). Hence, whereas there is mixed evidence on the autonomy of nudges in hypothetical settings (e.g., Michaelsen et al., 2021), nudges can likely be considered autonomy-preserving in real-life settings if the same nudges are considered autonomy-preserving in a hypothetical setting.

This study received ethical approval (Faculty's Ethical Review Committee of the Faculty of Law, Economics and Governance, Utrecht University; no. 2022-001). Data were collected via the Microsoft Office 365 portal of the organization. The organization signed a formal agreement to share and enable publication of anonymized data. The main data for this study are available at https://osf.io/6n2g4/?view_only=895be1c46d384867b52e22ff30892ba8 .

Participants

The quasi-field experiment was conducted at a large healthcare organization in the Netherlands. The organization has 22 locations in one city, employs around 2,300 employees (not including volunteers) and delivers care to more than 6,500 elderly clients.

Nudges

The nudges were identical to those in Study 1, but with a few minor changes for a better fit with the context. For the opinion leader nudge, rather than 'your HR manager', the name of the HR manager was included to increase the ecological validity of the nudge. Third, for the rule-of-thumb, the suggested 'within a day' communication option was a Teams message as this was the preferred mode of communication. The rule-of-thumb also included a brief note that any communication about clients should be done with a secure messaging tool.

Procedure

We measured email use (amount of sent emails) during eight consecutive weeks. All employees were subjected to the three nudges, distributed a week apart and starting in the second week. The nudges were distributed on three subsequent Mondays (28 March, 4 April and 11 April 2022) around 11:30 AM CEST to 3,038 work phones. The time was purposefully chosen because most employees experience a drop in daily workload after the morning duties. The SMS messages read 'Do you also want an emptier mailbox? Click here for a message/the second message/the last tips about emailing within [organization]' [Link]. Regards, [organization] Two weeks later (25 April, at 11:30 AM CEST), we posted the combination of all the nudges on the intranet of the organization.

To measure objective nudge effectiveness, the key dependent variable is the number of emails sent within the organization using administrative data available via Microsoft Office 365. In our main analysis, we excluded weekends and holidays as on these days, employees would email much less. We used linear mixed models to assess statistical significance, comparing each week to the week before in separate tests (Krueger and Tian, 2004) and using the Benjamin-Hochberg false discovery rate control to correct for multiple tests (Glickman et al., 2014). False discovery rate control is a less conservative alternative to the Bonferroni correction. It involves (a) sorting p-values in ascending order, (b) calculating the corrected p threshold per test by dividing the test number (e.g., 1 for test 1) by the total amount of tests (7 in this case) and multiplying this by the maximum false discovery rate (set at 0.05) and (c) declaring those tests with p-values lower than the corrected p threshold significant.

7.4. Results

7.4.1. Results Study 1: Survey experiment

Table 2 presents the correlations. One notable finding is that the score for perceived autonomy and all non-compliance estimates have significant negative correlations. This indicates perceived autonomy and subjective nudge effectiveness are positively correlated.

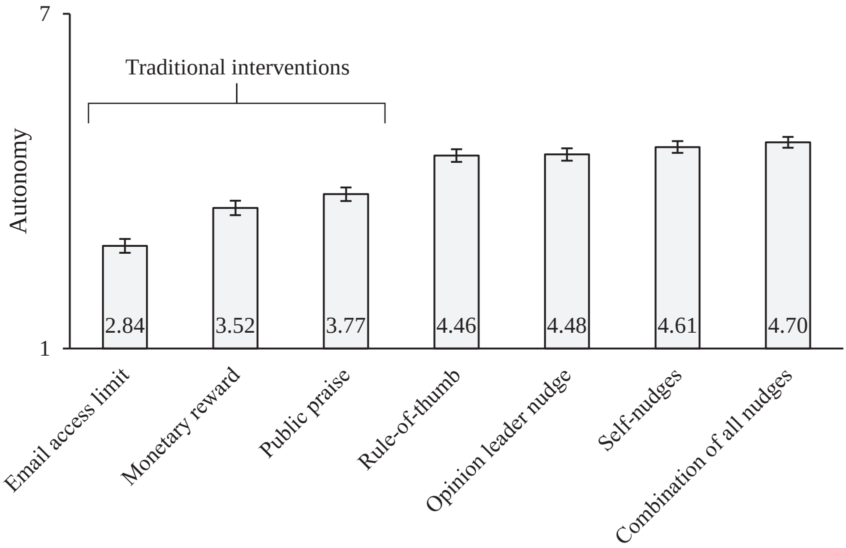
Table 2 *Correlations (n = 4,112)*

	M (SD)	1	2	3	4	5	6	7	8
1. Email volume	63.52 (86.25)	-	-	-	-	-	-	-	-
2. Email time	58.88 (66.98)	.54**	-	-	-	-	-	-	-
3. Email overload	3.06 (1.01)	.24**	.11**	-	-	-	-	-	-
4. Autonomy	4.05 (1.56)	-.03	-.03*	.002	-	-	-	-	-
5. Work engagement	3.88 (.71)	-.01	.01	-.17**	.07**	-	-	-	-
6. Self-admission rate	.68 (.47)	-.03	-.02	-.12**	-.34**	-.03*	-	-	-
7. Prevalence estimate	65.91 (25.25)	.11**	.08**	-.02	-.24**	-.08**	.43**	-	-
8. Admission estimate	60.2 (27.55)	.03	.03*	-.1**	-.18**	.02	.30**	.46**	-

Note. * $p < .05$, ** $p < .001$ (two-tailed). Correlations are Pearson except for those with email volume and email time, these are Spearman as for these variables the data indicated outliers. Variables 6-8 measure subjective nudge effectiveness. Note the self-admission rate is coded as a dummy (0: would comply, 1: would not comply). Appendix 3 provides more information on the measurement on the self-admission rates, prevalence estimates and admission estimates.

Figure 2 presents the means and 95% confidence intervals for perceived autonomy for each intervention.

Figure 2 *Nudges are seen as more autonomy-preserving than the midpoint (>4), and more autonomy- preserving than traditional interventions*



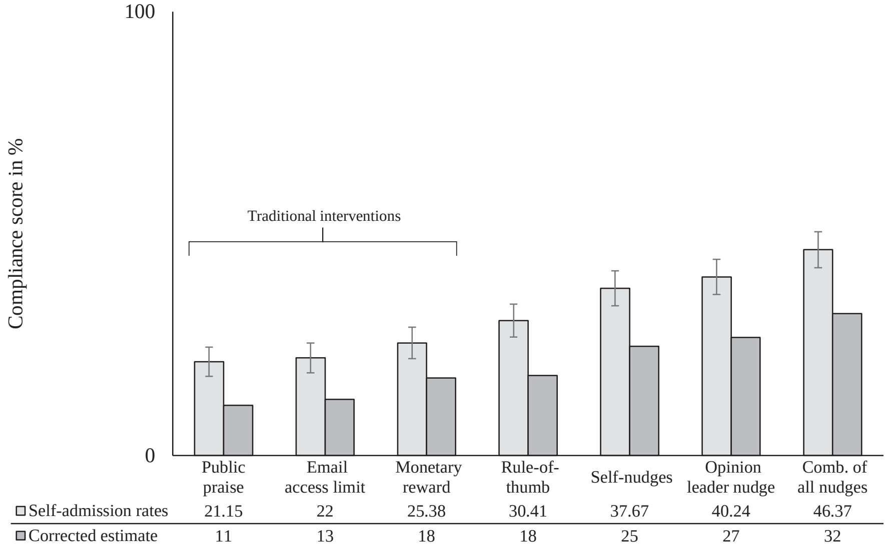
Note. Perceived autonomy scores show 95% confidence intervals.

A one-way analysis of variance showed that the effects on perceived autonomy differed significantly, $F(6, 1{,}823.18) = 135.51$, $p < 0.001$ ($\omega^2 = 0.17$)[2]. Post hoc analyses indicated that all traditional interventions scored significantly lower ($p < 0.001$) on perceived autonomy than any intervention with nudges. Besides, the email access limit scored significantly lower ($p < 0.001$) than the monetary reward and public praise. One significant difference between the nudges was found, the difference between all nudges and the rule-of-thumb is significant ($p = 0.028$). In sum, the results confirm our hypothesis that nudges are autonomy-preserving (scoring 4.46-4.7 on a 7-point scale), more so than traditional interventions (scoring 2.84-3.77).

Figure 3 presents the self-admission rates and the Bayesian Truth Serum results (recoded to self-admission rates of compliance to indicate how many respondents indicated they would send less emails) as a corrected conservative estimate of true compliance.

2 Welch ANOVA was reported and Games-Howell post hoc tests were used as equal variances could not be assumed, $F(6{,}4105) = 19.12$, $p < 0.001$.

Figure 3 *Nudges are perceived as more effective than the traditional interventions, but less than 50% of employees would comply with any intervention*



Note. Self-admission rates show 95% confidence intervals.

We conducted a one-way analysis of variance for the self-admission rates. This analysis showed that effects differed significantly, $F(6, 1{,}822.54) = 26.33$, $p < 0.001$ ($\omega^2 = 0.036$)[3]. Post hoc analyses indicated that all traditional interventions had significantly lower compliance than any nudge ($p < 0.001$), except for the rule-of-thumb. The rule-of-thumb had significantly higher compliance than the email access limit ($p = 0.017$) and public praise ($p = 0.005$), but not the monetary reward ($p = 0.465$). Besides, the rule-of-thumb had significantly lower compliance than the opinion leader nudge ($p = 0.007$) and all nudges ($p < 0.001$). Finally, the self-nudges had significantly lower compliance than the combination of all nudges ($p = 0.042$). The Bayesian truth serum indicates roughly the same, but more conservative, distribution, with a notable exception for the already insignificant difference between the monetary reward and the rule-of-thumb.

The results are in line with our main hypothesis. For a notable group of employees (30-46%), the nudges would be effective in reducing email use. The nudges are more effective in reducing email use than an email access limit, monetary reward or public praise (the latter scored 21-25% compliance) (except the difference between the rule-of-thumb and the monetary reward is non-significant).

3 Welch ANOVA was reported and Games-Howell post hoc tests were used as equal variances could not be assumed, $F(6,4105) = 87.21$, $p < 0.001$.

7.4.2. Results Study 2: quasi-field experiment

The first nudge was viewed 220 times, the second nudge 106 times and the third nudge 75 times. The combination of all nudges received 142 views in the first week (25 April-1 May) and 43 views in the second week (2 May-8 May), 185 views in total. This indicates many of the 3,038 recipients did not click on the link in the SMS. However, calculating a response rate on that total is misleading, as this does not equal the number of employees who use email. To estimate a more realistic response rate, we extracted email data on an individual level from the organization for the first week of the study (28 March-3 April). We first checked how many email IDs were in use that week (n = 2,618) and introduced the eligibility criterium of having sent at least one email that week, resulting in a total n of 1,189 active email users. This suggests that the estimated response rates for the first nudge was 18.50%, for the second nudge 8.92%, for the third nudge 6.30% and for the combination of all three nudges 15.56%.

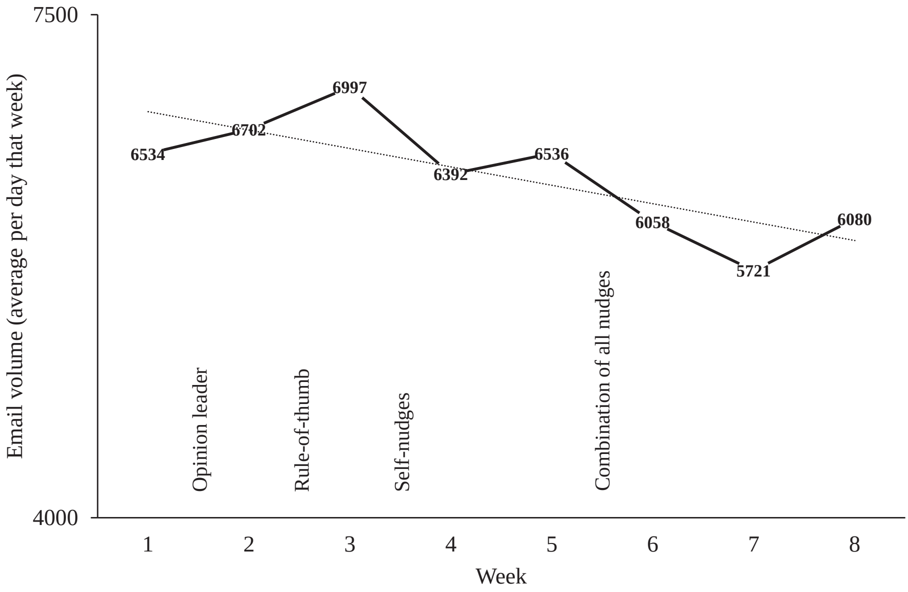
During the 8-week intervention period, a total of 236,785 emails were sent within the organization (excluding weekends and holidays). This is an average of 6,400 per day with a standard deviation of 1,125.

Figure 4 presents the average amount of email per day that was sent in the organization during each week of the intervention period, excluding holidays (Monday in week 5, Wednesday in week 6 and Thursday in week 7[4]) and weekends. We fit a linear trendline to indicate that, in general, email use decreased during the 8-week period ($y = -128.03x + 6953.6$; $R^2 = 0.593$). Between the first and last week of the quasi-experiment, average email use decreased by 6.95%. The biggest difference recorded was between week 3 and week 7 (–18.24%). Specifically, average email use decreased only during the week in which the self-nudges were distributed, and in the two weeks following the combination of all nudges.

7

4 Monday, April 18 was Easter Monday; Wednesday, April 27 was King's Day; Thursday, May 5 was Liberation Day.

Figure 4 *Email use decreases after the self-nudges and the combination of all nudges*



To test for statistical significance, we used seven linear mixed models to compare each week to the week before, using the variable describing the week as a repeated measure fixed factor, and the unstructured repeated covariance type. Table 3 presents the results of the separate linear mixed models and presents the corrected results using the Benjamin-Hochberg false discovery rate control to correct for multiple tests (Glickman et al., 2014).

Table 3 *Linear mixed models and false discovery rate control*

Test	Weeks	Days analyzed	Results	$p < .05$?	FDR p	$p <$ FDR p?
1	1 vs 2	M, Tu, W, Th, F	$t(4) = -1.67, p = .170$	No	0.04	No
2	2 vs 3	M, Tu, W, Th, F	$t(4) = -1.03, p = .360$	No	0.05	No
3	3 vs 4	M, Tu, W, Th, F	$t(4) = 2.76, p = .051$	No[a]	0.02	No
4	4 vs 5	Tu, W, Th, F	$t(4) = -5.93, p = .010$	Yes	0.007	No[a]
5	5 vs 6	Tu, Th, F	$t(4) = 6.30, p = .024$	Yes	0.01	No[a]
6	6 vs 7	M, Tu, F	$t(4) = 2.83, p = .106$	No	0.03	No
7	7 vs 8	M, Tu, W, F	$t(4) = -1.46, p = .240$	No	0.04	No

Note. In case any of the values of any day was missing (due to holidays), this day was removed from analysis in both weeks of a particular test. M = Monday; Tu = Tuesday; W = Wednesday; Th = Thursday; F = Friday. FDR p refers to the corrected p threshold calculated with false discovery rate control.[a] Indicates p-values very close to significance.

The results indicate that evidence for our hypothesis in the quasi-field experiment is mixed: email use did not decrease after presenting the opinion leader nudge or the rule-of-thumb, but it did decrease after presenting the self-nudges and the combination of all nudges. Statistical tests indicate a significant decrease ($p < 0.05$) after the combination of nudges, but this result is not significant after controlling for multiple tests. Across two months, however, the linear trendline indicates email use did generally decrease.

7.5. Discussion

7.5.1. Main findings

Our analysis of the survey experiment and quasi-field experiment yields three main findings. First, the nudges we developed were perceived as autonomy-preserving, and significantly more so than traditional interventions (email access limit, monetary reward and public praise). Second, our nudges were perceived as significantly more effective than the traditional interventions (except for the rule-of-thumb vs the monetary reward), but in general less than 50% of employees would comply with any intervention. We observe combining multiple nudges increased employees' perceptions of autonomy and effectiveness. We also found a positive correlation between perceived autonomy and subjective nudge effectiveness. Third, further evidence for the objective effectiveness of the nudges is presented in the quasi-field experiment. Email use in the healthcare organization decreased generally during the 8 weeks of our quasi-field experiment. Specific decreases were observed after the self-nudges and the combination of all nudges, albeit, after controlling for multiple tests, these effects did not reach conventional statistical significance levels.

7.5.2. Implications

The findings contribute to three major scholarly debates: nudge design, email use and interventions in the field.

Our paper contributes to the nudge literature by bringing nuance to the debates about the ability of nudges (1) to preserve autonomy (Hausman and Welch, 2010; Hansen and Jespersen, 2013), (2) to be effective in changing behavior, (Mertens et al., 2022; Szaszi et al., 2022; Maier et al., 2022) and (3) whether these two criticisms inherently present tensions (Wachner et al., 2020; Wachner et al., 2021; Mertens et al., 2022). We developed four scenarios for the autonomy and effectiveness of nudges. Using innovations in nudge design like self-nudges, we show that multiple types of nudges can be perceived as autonomy-preserving and effective in a survey setting. Another interesting finding is that combining nudges appears fruitful—suggesting that the sum may be perceived to be more than its parts. However, results are less pronounced in the field setting compared to the survey. There may be a variety of reasons, including that in the survey respondent's undivided attention is on a nudge, whereas in the field setting, employees may receive more messages simultaneously and choose not to engage (this is visible in the number of views the nudges received). Yet, ironically, the fact that a large share of employees chose not to engage does support the notion that nudges are autonomy-preserving even, and perhaps particularly, in field settings (Wachner et al., 2021). These results further emphasize the importance of the heterogeneity revolution (Bryan et al., 2021): rather than making statements of nudges in general, each nudge could have different consequences, different mechanisms and different effects depending on context.

Second, we contribute to the literature on email use by showing that nudges, and especially bundles of nudges, may help to reduce email use, which presents a serious threat to employee well-being (Brown et al., 2014; Reinke and Chamorro-Premuzic, 2014). We find nudges do have potential: employees are quite positive about them when it comes to autonomy and effectiveness, more so than policy instruments that are more commonly studied in the literature and used in organizations (Aguinis et al., 2013; Handgraaf et al., 2013; Tummers, 2019).

Third, our paper contributes methodologically, specifically on testing (behavioral) interventions. We introduced and redeveloped multiple ways in which nudges can be evaluated, prior to their implementation. In our pilot study, we assessed respondents' granular opinion on the nudges by distinguishing between feasibility, appropriateness, meaningfulness and effectiveness (Jordan et al., 2019). In our survey experiment, we used a modified version of the Bayesian truth serum to counter social desirability bias (Prelec, 2004; John et al., 2012). The results show that respondents are likely to overestimate their own compliance. At the same time, the truth serum tends to be conservative (John et al., 2012), meaning that the true value is likely to lie in between. This redeveloped serum could be a useful tool for scholars to evaluate nudges or other interventions to illicit more truthful responses. Finally, the comparative evaluation of nudges with traditional interventions has shed light on where nudges are positioned in the realm of policy and managerial interventions in general.

This research also has practical implications for managers and public policy. For managers, our research strengthens the argument that managers could turn to nudges as a valid and low-cost alternative to traditional policy instruments. Our findings suggest that unlike traditional policy instruments that might undermine employee autonomy (such as limiting email access), nudges can be autonomy-preserving and effective and can be used for concrete organizational challenges like email use. Our study also has implications for public policy. First, maintaining employee well-being in healthcare is an urgent public policy challenge in countries across the world. Well-being among healthcare employee has been increasingly put under pressure through, among others, the COVID-19 crisis (Spoorthy et al., 2020) and the aging workforce (Van Dalen et al., 2010). We have developed nudges to reduce a prevalent stressor in healthcare employees' jobs: email use. Studies have shown that email use can have very negative consequences for employees during and outside of work (Jackson et al., 2003; Kushlev and Dunn, 2015; Belkin et al., 2020; Giurge and Bohns, 2021). While it is unlikely that nudges will solve everything, we show how nudges can be part of efforts to contribute to improve the well-being of healthcare employee. Second, autonomy and effectiveness are critical issues in public policy. Scholars and practitioners have extensively debated whether nudges are a suitable policy instrument, including whether they are autonomy-preserving and effective (e.g., De Ridder et al., 2020; Tummers, 2022). Although our results may be context-dependent (e.g., Andersson and Almqvist, 2022; discussed below), our study suggests nudges can be autonomy-preserving and effective.

7.5.3. Limitations

We want to highlight several conceptual and methodological limitations. First, all nudges shared similarities: they were text-based, infographics, and sent via SMS or intranet. This allows for better comparison between the nudges because we minimize confounding variables. However, they do not fully represent the spectrum of what nudges can be (Thaler and Sunstein, 2021). We can therefore only draw conclusions from the nudges we tested. Compare, for example, the study by Andersson and Almqvist (2022), who find that the Swedish public prefers information and subsidies—both traditional policy instruments—above nudges. Following the logic of the heterogeneity revolution, future research should assess to what extent other types of nudges are able to preserve autonomy and be effective in other contexts (Bryan et al., 2021). Future research may also explore whether similar nudges, or bundles of nudges, could also be of help with different organizational challenges that affect well-being, like limiting work hours and maintaining a work-life balance (Pak et al., 2022). Finally, while we have compared nudges to traditional policy instruments across multiple studies, future research can explore potential causal mechanisms that explain why certain nudges have varying effects on autonomy and effectiveness. For example, in our survey experiment, the opinion leader nudge scored highest on individual effectiveness. It is possible that this nudge is more effective because it uses role modeling behaviors and fosters reciprocity between leaders and followers (e.g., Decuypere and Schaufeli, 2020).

Second, in this study, we conceptualized autonomy as the extent to whether nudges guarantee agency. However, Vugts et al. (2020) argue that nudges may also influence freedom of choice and self-constitution, which are the other two conceptualizations of autonomy. The question of whether a nudge is strengthening or empowering autonomy depends not only on the nudge itself, but also on the conceptualization of autonomy one focuses on (Vugts et al., 2020). More research is needed to better understand how nudges shape autonomy.

Third, our study presents both survey and quasi-field experimental evidence. The survey experiment does not measure actual behavior but intent. While intentions match behavior to some extent (e.g., Armitage and Conner, 2001; Hassan and Wright, 2020), a field experiment would introduce many aspects that a survey experiment lacks (Fishbein and Ajzen, 2010). The most important limitations of the quasi-experiment we used are the lack of a control group and randomized treatment allocation. Therefore, results may be biased by confounding variables (Shadish et al., 2002; Grant and Wall, 2009). Although we have taken several measures to deal with this, the effects of external events or elements unrelated to the treatment cannot be ruled out. Together with the limited amount of evidence in the quasi-field experiment, this constitutes a serious limitation. Also, the experimental period of 8 weeks is a considerable amount of time and a common timeline for work interventions (see Grant et al., 2014). Yet this does not warrant any claims about the true long-term effects of the nudges. In general, while some nudges, like defaults (e.g., Venema et al., 2018), can cause long-term effects, the long-term effects of nudges are insufficiently researched (Marchiori et al., 2017). Regarding the response rate, only a minority of employees viewed the nudges. Still, while future research could address these limitations by designing randomized controlled trials (Gerber and Green, 2012), our quasi-experimental approach does have benefits by testing nudges in the field and measuring effects on actual email use. We concur with Grant and Wall (2009), who argue a quasi-experiment can be a method of action research, providing an opportunity for collaboration between researchers and practitioners to jointly tackle organizational challenges.

7.5.4. Conclusion

This paper provides a nuanced perspective toward one of the most applied and debated behavioral interventions: nudges. Our theoretical approach and empirical substantiation indicate that nudges can be designed to be both autonomy-preserving and effective. Going forward, scholars and practitioners can leverage these insights to maximize the potential of what nudges can do.

Chapter 8

Concluding *Working on Well-being*

8.1. Introduction

In this dissertation, we studied healthcare employee well-being and how contemporary leadership approaches can contribute to it. Three main questions guided our study of *Working on Well-being:* (1) How can we deepen our understanding of employee well-being in healthcare? (2) How can leaders use empowerment to contribute to employee well-being in healthcare? (3) How can leaders use behavioral insights to contribute to employee well-being in healthcare? Below, we first provide a summary of the main conclusions. Next, we discuss the broader implications for theory, methods and practice, and discuss the limitations of the studies.

8.2. Main conclusions

RQ1: *How can we deepen our understanding of employee well-being in healthcare?*

Research showed that the well-being of healthcare employees is under pressure. Healthcare employees experience a range of increasing job demands, which, in combination with low job resources, may lead to burnout (Demerouti and Bakker, 2023; Patel et al., 2018). Burnout is associated with a variety of health problems and decreases in performance (e.g., Wen et al., 2016; Yang and Hayes, 2020). However, both causes and consequences of employee well-being in healthcare are complicated and multifaceted, which emphasizes the need to deepen our understanding of factors that affect well-being (Kniffin et al., 2021).

We defined two gaps in the literature on employee well-being that we aimed to address. First, we observed that during COVID-19, few studies paid attention to how this crisis affected groups of healthcare employees differently in the Dutch context (TNO et al., 2020). This is important as literature in other settings suggests that some groups of healthcare employees may be more at risk than others (Shreffler et al., 2020). Second, almost all studies that measure employee well-being do so with traditional, validated scales (Bakker et al., 2014). Such scales provide predetermined boundaries of our knowledge of employee well-being and limit the ability to make new theoretical discoveries (Balducci and Marinova, 2018).

Chapters 2 and 3 addressed those gaps and deepened our understanding of employee well-being. First, we made an empirical contribution in **chapter 2,** showing that healthcare employees report lower well-being when exposed to a threatening job demand: patients with a virus. By paying attention to risk groups of employees based on work and personal characteristics, we went beyond general patterns in employee well-being. This allows for tailored solutions that could be more effective in increasing well-being, such as job redesign to support elderly female employees (Kniffin et al., 2021).

Next, in **chapter 3,** we showed how innovative techniques such as text mining can improve our understanding of the multidimensionality of the employee well-being construct of work engagement. Our findings painted a richer picture of what it means to feel well at work than traditional survey measures (Balducci and Marinova, 2018; Kobayashi et al., 2018). Consequently, our study enabled critical assessment of the literature and explored how text mining can further contribute to studying employee well-being.

RQ2: *How can leaders use empowerment to contribute to employee well-being in healthcare?*

The literature on empowerment offered a first avenue for leadership that we aimed to explore: transferring influence from leaders to employees (Amundsen and Martinsen, 2014). We discussed two leadership styles that are key to such an approach: empowering and shared leadership. Empowering leadership is the process in which leaders provide employees with more autonomy and foster participation in decision-making (Ahearne et al., 2005), while shared leadership is a situation where leadership is dispersed among team members (Zhu et al., 2018). Scholars have explored how both styles impact employee well-being, showing that both empowering and shared leadership may increase it by providing job resources such as autonomy (Park et al., 2017) and collective efficacy (Zhu et al., 2018).

However, we showed that the literature on both leadership styles has urgent knowledge gaps. For empowering leadership, more recently, studies have found that empowering leadership may have dark sides (Cheong et al., 2016). Such findings conflict with the general literature (Kim et al., 2018; Lee et al., 2018) but may be partly explained by the neglect of one factor: how does context determine the effectiveness of empowering leadership (Kim et al, 2018; Sims et al., 2009)? Next, for shared leadership, studies have mainly focused on team and formal leader antecedents (Zhu et al., 2018). However, we knew little about employees' willingness to execute shared leadership, which limited realistic assessment of the promise of shared leadership: to what extent are employees willing to execute concrete shared leadership behaviors in organizations (e.g., Jønsson et al., 2016; Yukl, 2002)?

In chapters 4 and 5, we found that context and employee willingness are essential boundary conditions for the effectiveness of empowerment as a leadership approach. First, in **chapter 4,** we found that context matters in deciding whether empowering leadership is desirable (Sims Jr. et al., 2009). Specifically, empowering leadership can be a valuable strategy to increase employee well-being in crises with low intensity. However, empowering leadership was not effective in a high-intensity crisis. These findings helped explain mixed findings regarding empowering leadership (Cheong et al., 2016) and nuanced the potential of empowerment in crisis leadership ('t Hart and Tummers, 2019; Antonakis et al., 2003).

Secondly, in **chapter 5,** we found that employee willingness matters in evaluating the potential of shared leadership. We presented a bottom-up perspective of how shared leadership emergence is affected by the personal considerations of employees. We found that individual willingness is affected by leadership behavior, personal characteristics, and context. Finally, we discussed systemic issues that may help explain such individual preferences (e.g., Fleenor et al., 2010; Ryan et al., 2016).

RQ3: *How can leaders use behavioral insights to contribute to employee well-being in healthcare?*

The second avenue for leadership that we studied concerns behavioral insights. Behavioral insights are empirically verified insights from the behavioral sciences that center around the observation that our decision-making is not entirely rational nor optimal and highly influenced by our environment (e.g., Simon, 1955; Tversky and Kahneman, 1974). Based on such insights, interventions such as nudges can be designed to influence behaviors (Thaler and Sunstein, 2021).

While some studies have suggested that behavioral insights may help foster healthy behaviors or reduce stress (Georganta and Montgomery, 2016; Nagtegaal et al., 2019; Weintraub et al., 2021), there is little research on the potential of behavioral insights as part of a leadership approach to impact employee well-being positively. Besides, scholars have identified avenues to improve studies on behavioral insights within public administration generally. First, studies should go beyond quick wins to better understand the mechanisms of behavioral change more (Bhanot and Linos, 2020). Second, studies should critically assess the pros and cons of behavioral interventions like nudges (Bhanot and Linos, 2020) and design more (quasi) field experiments to assess them (Hassan and Wright, 2020).

Chapters 6 and 7 employed behavioral insights in the context of employee well-being-related behaviors: wrongdoings in organizations and email overuse impact employee well-being negatively (e.g., Brown et al., 2014; Searle and Rice, 2020). Besides, both chapters addressed the two avenues to understand the potential of behavioral insights better by paying attention to the mechanisms of behavior change and rigorous testing of interventions.

First, in **chapter 6,** we extended a recent experimental study that showed motivation activation could be a behavioral intervention to increase ethical reporting (Meyer-Sahling et al., 2019). We aimed to increase our understanding of the mechanisms of how motivation activation works (Münscher et al., 2016; Vogel and Willems, 2020). Through zooming in on some mechanisms of behavior change, we found that effectiveness varies depending on the dependent variable (i.e., who the wrongdoer is) and the type and reported level of the independent variable (i.e., which motivation and how high that motivation is). We show how such factors are important to consider when applying behavioral interventions in the field.

Second, in **chapter 7,** we aimed to test two returning criticisms of nudges: their inability to preserve autonomy (Hausman and Welch, 2010; Wilkinson, 2013) and their ineffectiveness (Maier et al., 2022; Mertens et al., 2022). We assessed nudges in both survey and quasi-field settings and found that our innovative nudges were perceived as both autonomy-preserving and effective in reducing email use. This indicates that nudges can be a suitable alternative to traditional policy interventions like protocols and that behavioral insights can provide leaders with additional tools to improve employee well-being.

8.3. Implications and future research suggestions

In this section, we first discuss theoretical and methodological implications of this thesis. Next, we formulate implications for practice.

8.3.1. Theoretical implications

On improving our understanding of employee well-being

We showed how our understanding of employee well-being can be improved when we, first, differentiate between groups of employees. **Chapter 2** 's main lesson concerns the negative impact that COVID-19 appeared to have on healthcare employee well-being.

These findings align with studies conducted in other settings (e.g., Shreffler et al., 2020) but shed light on the effects of COVID-19 within the Dutch context.

Yet how can we explain those findings? One explanation is that dealing with COVID-19 patients constitutes a threat demand: a job demand associated with increased distress and exhaustion (Tuckey et al., 2015). In this specific case, the threat demands include working under an elevated risk of infection and experiencing highly emotional labor. As such, a threat demand is distinct from hindrance or challenge demands—demands that block goals and decrease dedication or demands that foster gain and increase motivation—and it may also require a different solution. Perhaps the best solution to threat demands is to take away the threat or, if that is not possible, limit exposure to the threat as much as possible. This raises a normative question: should more vulnerable employees be less or not at all deployed in crises?

An alternative approach was advocated by Manzano García and Ayala Calvo (2021), who showed that the perceived threat of COVID-19 explained burnout among nurses. They suggested that organizations should ensure that emergencies are not experienced as a threat by, among others, providing sufficient resources and maintaining clear communication. A question for further research is to what extent resources to help employees deal with the crisis, such as respite rooms, counseling, and stress management, have been able to help reduce threat demands (Wei et al., 2020). What job resources are most promising in navigating the dangers and insecurities that employees face in crises (Demerouti and Bakker, 2023; Kniffin et al., 2021)?

The second avenue to improve our understanding of well-being is through innovating measurements. In **chapter 3,** we introduced text mining to study employee well-being, a literature mainly focused on traditional forms of analysis with validated scales. We showed that text mining adds multidimensionality and complexity, which allows us to confirm, extend and question theoretical frameworks (Balducci and Marinova, 2018; Kobayashi et al., 2021). Therefore, as an approach that combines the advantages of inductive and deductive, large-scale and context-rich research, text mining presents an alternative to traditional forms of analysis. However, scholars should bear in mind that text mining does come with its challenges: our attempts at classification were only moderately successful, and text mining requires a lot of expertise and access to substantial amounts of data to be carried out successfully.

Future research could finetune and expand on our text mining study. For example, our linguistic findings suggested that future research may investigate what other markers of employee well-being are visible in speech or writing. Such attempts have mostly been made within the context of clinical research (e.g., Franklin and Thompson, 2005) and not so much within well-being research.

Additionally, some findings that contrast earlier studies also beg for further research. For example, why do employees who are low-engaged refer to their managers more negatively? Managers can influence employee well-being in a positive or negative way (e.g., Barnes et al., 2015; Tummers and Bakker, 2021). The results suggest there must be a reason employees mention their managers more when they are a negative influence. One explanation is that positive leadership behaviors are seen more as self-evident (Toegel et al., 2013). However, we should also not underestimate the challenge that COVID-19 has posed for leadership in organizations (Graham and Woodhead, 2021). Either way, the findings suggest that managers are an important factor in employees' work engagement.

Another avenue for future research concerns the data that could be mined. Herein, scholars could look beyond surveys and use real-life data—such as emails, blog posts or meeting proceedings—to analyze how markers of employee well-being can be detected in the vast amount of unstructured data that organizations possess. Screening and identifying employees whose well-being is compromised at an early stage may help combat absenteeism and turnover (e.g., He, Veldkamp, & de Vries, 2012). What is more, some studies have shown that respondents may even prefer open response formats to rating scales when it comes to reporting mental health (e.g., Sikström et al., 2023). In cases where open questions are preferred for such reasons or simply because of time constraints, text mining allows efficient analysis of the rich data that respondents provide.

Finally, we see opportunities in the combination of text mining with more traditional research. For example, text mining could support the development of measurement scales by analyzing what words employees use to arrive at more ecologically valid scales (Kobayashi et al., 2021). Scale language can impact the response (see, e.g., studies on wording effects; Horan et al., 2003) and text mining may help reduce that bias.

The potential of empowerment as a leadership approach

Our findings showed that using empowerment as a leadership approach is unlikely to be universally effective. The primary implication of our studies on empowerment is that scholars should consider two boundary conditions: context and employee willingness.

First, context is important. We found that context determines the effectiveness of empowering leadership on employee well-being (**chapter 4**). Our study suggests that empowerment is not effective in a crisis for the energetic dimension of well-being.

This is in line with studies that show that the energetic dimension (i.e., vigor) is more prone to a health impairment process triggered by stressors than the motivational dimension (i.e., dedication) (Bakker and Demerouti, 2017). Our findings help explain why empowering leadership can have 'two faces' (Cheong et al., 2016) and nuance the more positive findings from recent meta-analyses (e.g., Lee et al., 2018).

There are multiple ways to expand research on the boundary conditions of empowerment. A major question remains whether empowering leadership is generally less effective in crises (Demerouti and Bakker, 2023). The literature is divided here. On the one hand, some scholars have argued that crises require bottom-up rather than top-down control to increase performance (Boin and 't Hart, 2003). This thought is also echoed in the literature on shared leadership: the need for shared leadership may be higher when organizations face turbulent times (Lund and Andersen, 2023). Such empowering strategies may also increase employee resilience (McDonald et al., 2016). On the other hand, scholars have argued that directive leadership may be more efficient in crises, as coordination is less ambiguous (e.g., Pearce et al., 2003). Likewise, crises are shown to increase directive leadership: the threat-rigidity hypothesis describes the phenomenon that leaders tend to switch to more directive leadership styles after a crisis (Stoker et al., 2019).

So, is empowerment or directive leadership the answer? The answer seems to be that it depends. A recent study showed that participative or directive leadership was more effective depending on whether the crisis was familiar or not: directive leadership was more effective in familiar emergencies, but participative leadership was more effective in unfamiliar emergencies (Post et al., 2022). Likewise, a leader's place in the organizational hierarchy may affect what behaviours are fitting (Schmidt and Groeneveld, 2021). Such insights fit the principle of situational leadership: scholars have since long argued that leadership approaches are often more or less effective depending on the specific situation (e.g., Sims Jr. et al., 2009; Stoker et al., 2019). Even more broadly, insights from the social sciences are rarely universally applicable. Instead, knowledge is localized. This heterogeneity should not be ignored but studied (Bryan et al., 2021). Our findings indicate that leaders should take multiple factors into account when considering empowerment, such as the type of crisis and the outcome for which they are aiming.

The second boundary condition that leaders should consider is employee willingness. We found that employees' willingness to execute shared leadership varies based on the specific leadership behavior, personal characteristics, and context (**chapter 5**). This is a novel contribution as much of the literature has conceptualized shared leadership more generically (e.g., Carson et al., 2007) and focused on higher-level antecedents (e.g., Wu et al., 2020). Our findings on shared leadership shed light on its potential by conceptualizing it as a collection of specialized behaviors and assessing whether individual employees would want to execute those behaviors (Yukl, 2002).

The many findings of this study stimulate further exploration of employee willingness to execute shared leadership. One exciting avenue is that scholars could critically look at the validity of self-reporting on shared leadership. Specifically, reported preferences are likely a product of both true preferences and biases. In our study, we might have seen this in employees' preferences regarding the effort that leadership behaviors take: employees preferred behaviors that take less time per week, but they did not mind how long the leadership behavior lasts. The former finding arguably represents a real preference as increases in workload are known to affect well-being negatively and employees may want to avoid that (Demerouti and Bakker, 2017). Nevertheless, the lack of effects when it comes to leadership longevity may well be due to presentism bias: people underestimate the impact of events in the future, such as how their preferences may change (Bauckham et al., 2019). This raises various questions for further research on how self-reported preferences may deviate from actual practices. For example, we adopted a more formal approach to shared leadership because employees were asked to perform a task. However, employees may also more informally take up shared leadership behaviors without even realizing it (Carson et al., 2007). What behaviors are or are not seen as leadership behaviors? Moreover, to what extent do self-reported preferences and unconscious leadership behaviors overlap? Likewise, we know that self- and other ratings of leadership capabilities differ (Fleenor et al., 2010). But to what extent does willingness coincide with the right competencies when it comes to shared leadership?

The potential of behavioral insights as a leadership approach

Besides empowerment, leaders could employ behavioral insights to improve employee well-being. In this dissertation, we showed two ways in which scholars could assess the potential of behavioral insights more rigorously: by exploring the mechanisms of behavior change for behavioral interventions and by testing behavioral interventions elaborately in a field setting (Bhanot and Linos, 2020; Hassan and Wright, 2020).

Assessing the potential of motivation activation for willingness to report wrongdoings rigorously in **chapter 6,** we found how effectiveness varies depending on the type and level of motivation and the wrongdoer. Doing so, we extended the seminal study by Meyer-Sahling et al. (2019) and contributed theoretically to the literature on ethical behavior and motivation (e.g., Olsen et al., 2019; Ripoll, 2019).

Our study on motivation activation raises several questions that scholars could address. Testing effects in field settings is an important one, but the mechanisms that we explored also beg for further inquiry. For example, activating motivation only works for reporting patients but not for peers. Why? The relationship between the reporter and the wrongdoer appears important (e.g., Graaf, 2010). Various concepts may explain this relationship, and future research may test which plays a role. For example, the literature on professionalism suggests that employees may be collectively self-interested, aiming for a monopoly on the provision of services (Andersen and Petersen, 2012). Reporting peers goes against that principle. Additionally, in social identity theory scholars have discussed the phenomenon of ingroup favoritism (Hogg, 2018). Healthcare employees may see peers, but not patients, as their in-group, for whom other rules apply when it comes to handling failures.

Another avenue that scholars may pursue is that, in contrast to Meyer-Sahling et al. (2019), we found ceiling effects: highly motivated employees could not be activated. Scholars may look further into the existence and causes of such ceiling or floor effects and what they mean for the effectiveness of behavioral interventions. We expect that activation is generally most adequate for the group that scores average: there must be a baseline of motivation to activate, but there must also be room for improvement.

Next, developing and testing innovative nudges across a survey and quasi-field experiment in **chapter 7** showed that nudges can be autonomy-preserving and effective tools to reduce email use. With this, we provided nuance to the debates on nudge effectiveness (Mertens et al., 2022; Szaszi et al., 2022) and autonomy (Wachner et al., 2021; Wilkinson, 2013).

The fact that employees believe that nudges preserve autonomy is an important finding. Nudges, like any policy instrument, are nothing without legitimacy and the support of employees themselves (Tummers, 2019). However, the literature also includes contrasting evidence (Andersson and Almqvist, 2022). This begs the question: what types of nudges can preserve autonomy and what types are not? Here, overview studies like Münscher et al. (2016) may help discuss the wide variety of available nudges. Besides, our findings place an important reservation on the perceived autonomy of more traditional policy instruments that are commonly studied and used in organizations: they may not be perceived very positively by employees (Aguinis et al., 2013; Handgraaf et al., 2013; Tummers, 2019).

However, we also argued that outcomes depend on the conceptualization of autonomy itself (Vugts et al., 2020). Different conceptualizations cloud the current literature, often used implicitly. We interpreted autonomy as promoting agency. This is a higher threshold for nudges than maintaining freedom of choice, which was the initial interpretation of autonomy (Thaler and Sunstein, 2008). Besides, nudges could perhaps simultaneously increase autonomy according to one interpretation, and decrease autonomy according to another (Vugts et al., 2020). For example, a nudge could limit choices but at the same time help someone reach their goals. The nudge debate on autonomy is more complex than it appears and future research should do justice to its complexity.

Finally, we observed that behavioral insights are scarcely studied in organizational contexts to improve well-being. Therefore, we should study how we can employ behavioral insights to foster other positive behaviors, such as stimulating work-life balance, reducing telepressure (Barber en Santuzzi, 2015), and fostering healthy workplace behaviors like taking breaks and setting boundaries. By working together with practitioners, experimental researchers could execute a form of action research, jointly tackling challenges that contemporary organizations face (Grant and Wall, 2009).

Combining empowerment and behavioral insights

We have addressed the potential of empowerment and behavioral insights separately. Going forward, scholars could study how empowerment and behavioral insights interact when employed jointly and how one can strengthen the other. For example, does using behavioral insights for one goal leave employees with more cognitive and motivational abilities to be empowered for another goal? Such combined effects could be explored. Additionally, behavioral insights could be employed to stimulate empowerment. Can leaders leverage behavioral insights to make it easier for employees to display shared leadership? Finally, can employees be empowered to make more use of behavioral insights? Herein, self-nudges offer a way in which employees could actively change decision environments for themselves and their colleagues (Reijula and Hertwig, 2022).

Admittedly, a leadership approach that includes both empowerment and behavioral insights may have paradoxical implications. For example, could nudges counteract feeling empowered? Such tendencies could be explored (see, e.g., Yang et al., 2021; Zhang et al., 2015). In the leadership literature, scholars recently called for more theoretical integration with other disciplines, such as psychology (Antonakis et al., 2023). In the literature on behavioral insights, scholars have urged us to link behavioral insights to more systemic and complex issues (Bhanot and Linos, 2020). Studying empowerment and behavioral insights jointly is one way to do so.

A note on individual preferences and systemic causes of inequality

A final implication follows from the findings throughout this dissertation. In the studies, we found that multiple personal or work characteristics affect one's well-being or attitude toward leadership. Such findings may have deeper-rooted causes that should not be ignored. Take gender. In healthcare, most employees are female, but leaders are relatively often male (Mousa et al., 2021). Healthcare organizations sometimes show hierarchical and patriarchal tendencies that foster an environment of systemic inequality between men and women (WHO, 2019). This may very well explain some of the findings in this research. Women report more health problems than men (chapter 2). This is likely connected to the fact that women more often work in jobs with direct patient contact. Women are less willing to execute shared leadership (chapter 5). This links to the literature showing how men and women perceive leadership differently: leadership may be regarded as a masculine activity (Elliott and Stead, 2008) and men may romanticize leadership to a greater extent (Collinson et al., 2018). We have discussed the potential systemic causes, like socialized norms, which hamper women's advancement in healthcare organizations. At the same time, our understanding of the stereotypes, sexism and functioning of hierarchies within healthcare contexts is limited (Ryan et al., 2016). In the future, scholars may study such underlying causes to gain a better understanding of why and how inequality proliferates and how solutions that we advocate may or may not help. In doing so, they should also go beyond the male/female dichotomy and pay attention to the often-overlooked cases of healthcare employees who report a gender other than male or female.

8.3.2. Methodological implications

Throughout the chapters of this dissertation, we aimed to make two methodological contributions that could help scholars design studies in the future. First, we tested many solutions to innovative survey designs and second, we adopted Open Science practices.

Survey design solutions

Throughout this dissertation, we tested survey design solutions to improve the quality of causal claims. In public administration, surveys have become an increasingly common tool to collect substantial amounts of data from respondents systematically and efficiently, from opinions and attitudes to factual knowledge, and generalize these to a larger population (Groeneveld et al., 2015; Swidorski, 1980). At the same time, surveys can be limited in the extent to which they can address causal relationships due to endogeneity. Endogeneity is the result of, among others, the existence of uncontrolled confounding variables or simultaneity between two variables of interest (Podsakoff et al., 2003). One solution is to design experiments (Bouwman and Grimmelikhuijsen, 2016). The golden standard in experimentation is the randomized controlled trial (RCT), in which a randomly selected group of participants receives a treatment. In contrast, another group functions as the control group (Gerber and Green, 2012). While these experiments can offer precious knowledge, they are also notoriously hard to execute, especially in the field (Hansen and Tummers, 2020). Another way to go is to consider how surveys could be improved. In this dissertation, the solutions to improve survey designs used are divided into three categories. Table 1 presents an overview and indicates in which chapters the innovations were applied. In each of the chapters, the methods were elaborately discussed. Below, we discuss all of them briefly.

Table 1 *Survey design solutions*

Solution	Specific method	Chapter
Temporality	Repeated cross-sectional surveys	3
	Longitudinal survey with self-generated identification codes	4
Teaming	Combine survey variables with administrative data	4
	Combine survey with quasi-field experiment	7
Techniques	Conjoint experiment	5
	Question order design	6
	Text mining (classification)	3
	Bayesian truth serum	7

First, adding temporality can be helpful as it is one of the preconditions for causal claims (Shadish et al., 2002). We employed a repeated cross-sectional design and a longitudinal design to add temporality. For the longitudinal design, we used self-generated identification codes (Schnell et al., 2010). Self-generated identification codes offer a way to link respondents' data across survey waves anonymously by having them generate a code based on questions of which the answers are known to the respondent but unknown to the researchers (e.g., what is the first letter of the place you are born in?).

Second, combining surveys with other data sources generates a more robust design. We showed that by combining survey variables with administrative data, we can effectively design a natural experiment (Sieweke and Santoni, 2020). Besides, we combined a survey experiment with a quasi-field experiment to assess both perceptions and behaviors.

Third, using new techniques in survey designs or measures allowed us to address some of the shortcomings surveys have. For example, survey experimentation allows the test of causal claims but offers a more accessible and increasingly used alternative to field experiments, although one should carefully consider the external validity of such designs (James et al., 2017; Jilke and Van Ryzin, 2017).

We employed two survey experimental designs. We used a conjoint experiment (Hainmueller et al., 2014), a design adopted from mathematical psychology that allows us to experimentally manipulate variables to assess their independent effects (Karren and Barringer, 2002). We also used a question order design (Meyer-Sahling et al., 2019), a method that allows us to assess causal effects by using randomized question orders.

Finally, we assessed two innovative survey measures. We examined the benefits of text mining in surveys. Text mining allows to study unstructured text on a large scale. In combination with traditional work engagement scores, we classified employees as having high or low work engagement based on the narratives they write and we interpreted the text features that contributed to correct classification (Balducci and Marinova, 2018). Next, we used a Bayesian truth serum to counter social desirability bias (Prelec, 2004). The serum is a scoring algorithm that combines the personal answers of respondents with estimates of others, presenting a combined estimate that increases credibility and reduces social desirability bias (John et al., 2012, p. 526).

The above solutions provide a toolbox of methods that scholars could apply to innovate survey designs. Going forward, this toolbox could be discussed in terms of its ability to improve the quality of survey designs, and scholars could expand the toolbox by testing other solutions or testing the solutions in other settings to assess their external validity.

Open Science practices

The second methodological contribution is that we used Open Science practices. Open Science refers to 'an array of practices that promote openness, integrity, and reproducibility in research' (Banks et al., 2018, p.111), emerging in response to the 'reproducibility crisis' (Baker, 2016). This crisis, referring to the failure to reproduce studies, was caused partly by questionable research practices that are prevalent in the scientific community (Open Science Collaboration, 2015). Examples of such practices are selective reporting, *p*-hacking or HARKing (Hypothesizing After the Results are Known) (John et al., 2012; Kerr, 1998). The purpose of Open Science is to prevent questionable research practices or outright research misconduct by advocating for practices like preregistration, open-access publishing, open data, and replication (Banks et al., 2018; Nosek, 2018). Only recently have Open Science practices gained traction with public administration research (e.g., Pedersen and Strich, 2018). A 2020 review of field experiments in public administration found that only two out of 42 studies were preregistered (Hansen and Tummers, 2020). Vogel and Xu (2021) conclude that the discipline of public administration has not adopted preregistration widely yet. In this dissertation, we aimed to show how to design studies that adhere to Open

Science practices as much as possible. Table 2 presents an overview of these practices and in which chapters they were applied.

Table 2 *Open Science practices*

Practice	Explanation	Chapter
Ethical review	Research design is assessed by an ethical review committee	2-7
Preregistration	Research hypotheses and analysis plan are submitted to an online repository prior to execution of the research	3-7
Open access	Scientific publication is freely accessible to anyone	2-7
Open syntax	Syntax for the research is provided alongside scientific publication	3, 5, 7
Open data	Data from the research are provided alongside scientific publication	2, 4*, 5, 7
Open research materials	Additional research materials are provided alongside scientific publication	3-7

Note. Not all practices were applicable in every chapter. *Data available upon request.

Going forward, we urge scholars to adhere to Open Science practices because they offer a solution to avoiding questionable research practices (Banks et al., 2018). However, adhering to all Open Science practices is easier said than done. Open Science requires thinking in advance. Many of the good practices should be planned and implemented before any of the empirical research starts. For example, obtaining ethical review and writing a preregistration should now become the first steps in the empirical research process and this requires time. In adhering to Open Science, this research has been a learning experience. We found out, for example, that we could not share data for every study openly if legal agreements had already been made. Besides, we should acknowledge that we are privileged to have a government that prioritizes open-access publishing (see openaccess.nl). Unfortunately, not everyone has access to those resources. Nevertheless, we believe that through trial and error, the Open Science practices can and should become both internalized norms and external requirements for scientific research.

8.3.3. Implications for practice

The insights that managers and policy makers can take from this dissertation are twofold. We provide valuable knowledge that will allow them to *understand* employee well-being better and *improve* it.

First, it is vital that leaders genuinely *understand* how their employees are doing. Rich insights into healthcare employee well-being, the role of personal (e.g., gender and age) and work characteristics (e.g., job function), and the experiences that employees have are necessary. This is where practitioners need to execute research to get to know their organizations. The methods used in this research offer solutions that managers and policymakers could employ to study their employees' well-being while minimizing the additional burden of applied research on employees. Take employee satisfaction surveys. When leaders want to assess employees' preferences on a topic, using a conjoint rather than a vignette experiment is beneficial because not all scenarios have to be tested to draw conclusions (Hainmueller et al., 2014). Additionally, some research goals may be achieved by using text mining instead of validated scales so that employees could provide a short story instead of having to participate in extensive surveys (Balducci & Marinova, 2018). Further, studies could also take advantage of existing unstructured data in organizations, such as (anonymized) emails, messages or tweets. Finally, the Bayesian truth serum that we used may also inspire leaders to find ways to collect the actual attitudes of employees rather than leaders being told what they want to hear (Prelec, 2004).

Second, this dissertation helps leaders find innovative ways to *improve* employee well-being. We have discussed two leadership approaches. In some cases, empowerment may be the way to go: giving employees more autonomy and participation in decision-making can be a job resource that increases well-being. This can be done by, for example, making employees responsible for changing the ways of working in a team. In other cases, behavioral insights may be helpful: changing employees' environment with behavioral interventions so that healthy behaviors become more accessible to practice. Leaders could, for example, display opinion leadership to foster healthy workplace behaviors. In addition, we also discussed how both approaches might be used together: employees could be nudged to empowerment or empowered to nudge themselves and others. Table 3 summarizes examples of practices regarding understanding and improving employee well-being that leaders could explore.

Table 3 *Examples of practices for leaders*

Goal	Suggestions for practices	Chapter
Understand employee well-being better	Use text mining to create a richer picture of how employees feel at work	3
	Use a Bayesian truth serum to reduce social desirability bias in employees' responses	7
Improve employee well-being with empowerment	Make employees responsible for a project to change the way of working, e.g., by finding ways to decrease administrative burden	5
	Make employees responsible for improving relationships within the team, e.g., by becoming the team counsellor	5
Improve employee well-being with behavioral insights	Be an opinion leader, use yourself as a social reference point by setting an example of what healthy work behaviors entail	7
	Help employees boost self-control by giving them self-nudges for healthy behaviors, e.g., help employees reminding themselves to delay responses to email less	7

However, we also found boundary conditions for empowerment: leaders should consider whether the context is right and whether employees agree with the approach. There are a variety of contexts in which empowerment may fail. We showed that empowering leadership is not preferable in a severe crisis. Besides, employees' willingness to execute shared leadership depends on several factors: what leadership behavior is shared, the gender and age of the employee, and context. In this regard, it is vital to not only consider those factors themselves but also pay attention to underlying systemic causes. For example, sexism within the organization may be the cause of female employees being less willing to execute shared leadership (Ryan et al., 2016).

Likewise, behavioral insights and interventions are unlikely to work universally. Instead, such interventions are to be evaluated in a variety of contexts to understand fully how, when and why they work (Bryan et al., 2021). Whether a specific behavioral intervention works in a specific context can only be known when this is tested. Besides, employees' approval of behavioral interventions should be seen as a crucial part of making sure that behavioral interventions are legitimate (Tummers, 2019).

In sum, *Working on Well-being* means employee well-being is understood better and improved with appropriate use of empowerment and behavioral insights.

8.4. Limitations

In addition to limitations that have been discussed in the specific chapters, here we discuss general limitations of this dissertation.

First, there are methodological limitations to the survey studies in this dissertation. Surveys allow the study of large amounts of cases cost-effectively, but they also depend on self-reported measures. Hence, potential bias creeps in when respondents provide the answers they think are socially desirable (Grimmelikhuijsen et al., 2017). We cannot completely rule out such effects. In some chapters, we evaluated the likelihood of social desirability bias. We found there are good arguments to argue that this bias is not prevalent (chapters 5 and 6), and in chapter 7, we tested an approach to counter it (the Bayesian truth serum, John et al., 2012). Besides this general drawback, there are some specific drawbacks. Chapter 2 presented a cross-sectional survey. The results of this study should be interpreted as correlational (i.e., a statistical association) and not causal (i.e., a change in one variable leads to a change in another variable). For example, we found that healthcare workers who work with COVID-19 patients are more physically exhausted, but we cannot conclude that working with COVID-19 patients leads to more physical exhaustion. In other studies, we took measures to improve cause-and-effect claims, like using a repeated cross-sectional design (chapter 3), designing a natural experiment (chapter 4), or using survey experiments (chapters 5 and 6). However, such studies (chapters 2-4) still risk common source bias as they use the same source to collect multiple (i.e., independent and dependent) variables. At the same time, we have taken some measures to reduce the risk by comparing two survey waves (chapter 3) or using a fixed-effects analysis (chapter 4; Jakobsen and Jensen, 2015).

Next to methodological limitations, constraints apply to the samples. We found that our samples are generally representative of the general population of healthcare employees in the Netherlands. However, there are some exceptions. We collected respondents mainly from members of Stichting IZZ, a healthcare collective. This caused our samples to be overrepresented by older employees (older employees tend to participate in collective healthcare insurance more often than younger employees) and employees working in hospitals. Across chapters, we assessed representativeness by comparing sample characteristics to the general population (chapters 2-5) or the population of IZZ members (chapter 6). Another issue is that between 82% and 87% of our samples consisted of female employees, which is similar to the population of Dutch healthcare employees but does limit generalization to other sectors. More generally, all samples were taken from the general population of members of IZZ (the exception is chapter 7, which included a pilot survey among the general working population and a quasi-field experiment in a specific healthcare organization). Finally, a note on sample sizes. Most of our samples were large and provided sufficient power to test our hypotheses. Some scholars have suggested that studies can also be overpowered, in which case statistically significant effects can lose scientific significance (Zhang and Hughes, 2020). To limit that risk, we have carefully interpreted our results, using both significance levels and effect sizes.

Third, we also want to highlight limitations related to the topics and contexts studied. When studying the potential of empowerment or behavioral insights, we could only address a few of the knowledge gaps. For example, we have addressed the effects of empowering leadership in the COVID-19 context (chapter 4), analyzed the willingness of employees to execute a few types of shared leadership (chapter 5), and used behavioral insights to influence ethical reporting and email behaviors (chapters 6 and 7). Through choosing these topics, we have addressed a breadth of possibilities for *Working on Well-being,* but we acknowledge the limited scope of our topics and contexts. Besides, the COVID-19 crisis was a contextual variable in most of the studies. Yet while this crisis presented unique circumstances in its impact on society, high job demands and work stress were typical among healthcare employees before the crisis (Bakker et al., 2014). As such, the crisis may have amplified existing effects but is unlikely to limit generalization to its specific context.

Finally, we should note that almost all studies used a deductive, quantitative research strategy (except for the inductive study in chapter 3 and the qualitative pre-study in chapter 7). Any research strategy is conducted within the context of implicit or explicit ontological (i.e., how is reality construed) and epistemological (i.e., how can we study reality) beliefs (Haverland and Yanow, 2012; Raadschelders, 2011). The view of reality in quantitative research is often objectivist: social phenomena like organizations or leadership are seen as external entities. Consequentially, they tend to be studied using a positivist paradigm originating in the natural sciences: a process of developing theories and empirical testing of hypotheses (Raadschelders and Lee, 2011). To the extent that the studies in this dissertation breathed such a paradigm, they also shared its blind spots. For example, our methodologies and methods may have underexposed how employees subjectively interpret and give meaning to concepts like leadership or well-being (Hay, 2011).

8.5. Concluding *Working on Well-being*

In this dissertation, we have studied healthcare employee well-being and how leadership can improve it. We deepened our understanding of employee well-being itself, and we explored two contemporary approaches towards leadership that may improve well-being: empowerment and behavioral insights. The insights in this dissertation will hopefully make healthcare employment healthier. The well-being of patients or clients in healthcare organizations is unthinkable without appropriate care for the well-being of employees themselves.

Summary

Introduction

Healthcare employees are invaluable assets to healthcare organizations. In the Netherlands, their work contributes to a comparatively well-functioning healthcare system and good population health. However, employees' own well-being has increasingly been challenged. An accumulation of job demands, like long working hours and high emotional burden, is associated with higher risks of burnout. As a result, healthcare employees are more likely to develop occupational diseases like musculoskeletal disorders and have higher absenteeism. Sometimes, healthcare employees even consider leaving healthcare. How can we take care of healthcare employees? In this thesis, we aimed to deepen our understanding of employee well-being and develop strategies that could help improve well-being.

The literature has identified various job resources that positively impact well-being. One key resource is leadership. Research shows that leaders can improve employee well-being by influencing employees' job demands and resources. Traditionally, leadership is defined as a process of social influence toward goals that people in formal leadership roles enact toward their followers in an organization. In recent years, scholars have developed novel approaches towards leadership. In this dissertation, we studied two of those approaches: empowerment and behavioral insights.

First, we studied empowerment. According to some scholars, employees have become more skilled and knowledgeable than ever before. This is why they argued for the emancipation of the workforce. Leaders should step back and empower employees at work. This way, leadership becomes a joint venture of leaders and employees. Empowering employees often has positive effects on the well-being of both leaders and employees. However, some scholars pointed out that it can also have a dark side through, for example, increasing the workload. A concrete example of empowerment is that employees could execute a leadership behavior aimed at relations by, for example, becoming the team counsellor.

Secondly, we evaluated the potential of behavioral insights. This is based on the notion that people have limited decision-making capabilities. In our decisions, bounded rationality causes us to make predictable errors. In response, scholars have developed theories and interventions based on empirical evidence of how human cognition works. Leaders could use behavioral insights to stimulate employees to make decisions that improve their well-being. An example is that leaders could communicate that they are going to email less, which nudges employees to follow.

Research questions

We aimed to advance our understanding employee well-being in healthcare and how leadership can contribute to it by developing studies that address a variety of topics and display a breadth of possible approaches and tools. Three main research questions guided this process.

RQ1: *How can we deepen our understanding of employee well-being in healthcare?*

In chapters 2 and 3, we addressed two literature gaps on employee well-being by differentiating between groups of healthcare employees and by innovating the measurement of well-being. In **chapter 2,** we analyzed how the experiences of Dutch healthcare employees varied during the COVID-19 crisis. Furthermore, in **chapter 3,** we presented text mining as an innovative method to analyze employee well-being.

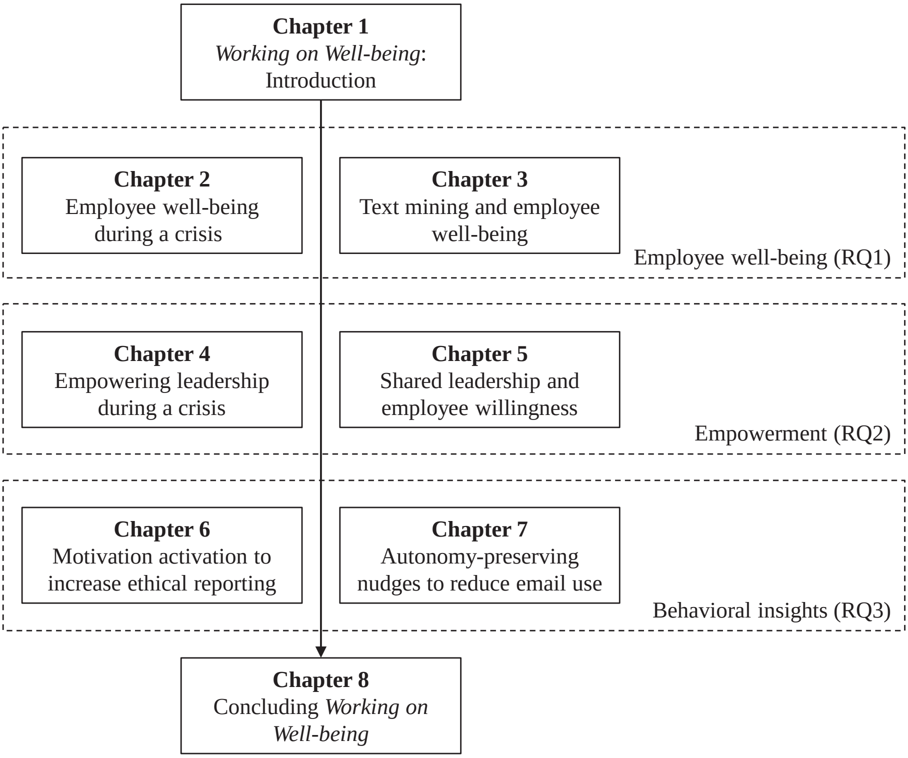
RQ2: *How can leaders use empowerment to contribute to employee well-being in healthcare?*

In chapters 4 and 5, we studied two leadership styles that are related to empowerment: empowering leadership and shared leadership. The literature mainly describes the positive effects of these styles, but there are limitations: few scholars have studied the role of context and the willingness of employees themselves. In **chapter 4,** we studied whether empowering leadership also has a positive effect on employee well-being in a specific context: during a crisis. Moreover, in **chapter 5,** we investigated which factors influence the individual willingness of employees to exercise shared leadership.

RQ3: *How can leaders use behavioral insights to contribute to employee well-being in healthcare?*

In chapters 6 and 7, we explored the potential of behavioral insights. To do so, scholars have suggested that we need to look closely at the mechanisms of behavior change and test interventions rigorously and in field settings. We addressed those gaps in the context of behaviors that are related to employee well-being. In **chapter 6,** we used a recent study that showed that one can activate the motivation of employees to increase ethical reporting. We tested whether this depends on the type and level of motivation and who the wrongdoer is. In **chapter 7,** we tested whether innovative nudges could be perceived as autonomy-preserving and reduce email use among healthcare employees. Figure 1 presents an overview of this thesis.

Figure 1 *Overview*



Chapter 2: Employee well-being during a crisis

In a survey study of 7,208 healthcare employees, we studied whether healthcare employees who worked with COVID-19 patients experienced lower well-being than other healthcare employees. We also studied what personal and work characteristics accounted for differences within the group of healthcare employees working with COVID-19 patients. We found that healthcare employees in direct contact with COVID-19 patients reported more sleep problems and more physical exhaustion. Besides, within that group of healthcare employees, those that are female, living alone, without leadership role or sufficient protective equipment reported lower well-being. While physical exhaustion was higher among older healthcare employees, mental exhaustion was higher among younger employees. These results emphasize the importance of adequate support for healthcare employees and argue for taking into account the specific circumstances of the individual healthcare employee.

Chapter 3: Text mining and employee well-being

In this chapter, we introduced text mining as a method to study employee well-being. We used two surveys (among 5,591 and 4,470 employees) to classify healthcare employees' self-written narratives into high or low work engagement using text mining. We found that psychological features—like positive and negative emotions—have 60% accuracy in classifying. Analyzing the text features that contribute to classification allowed us to deepen our understanding of work engagement by validating, extending or questioning the literature. For example, we extended the literature by showing that work engagement has a linguistic component: high-engaged employees more often used the first-person plural ('we'). Based on the results, we discussed how future research may use text mining to understand employee well-being better.

Chapter 4: Empowering leadership during a crisis

In this study, we combined a longitudinal survey (n = 468) with administrative data on COVID-19 hospitalization rates across provinces in a difference-in-differences natural experiment. We investigated whether the positive effect of empowering leadership on healthcare employee well-being would be present during the crisis. We found that for the motivational and cognitive dimensions of well-being, the effect of empowering leadership does not vary during a crisis. However, for the energetic dimension (e.g., mental exhaustion), empowering leadership is less effective in a crisis and can even harm the well-being of employees. These findings show that empowering leadership can have adverse effects in specific contexts.

Chapter 5: Shared leadership and employee willingness

In a conjoint experiment among 6,742 healthcare employees, we asked healthcare employees to evaluate leadership behaviors to execute in a shared leadership role. We found that a notable share of employees was willing to execute shared leadership. We also found that willingness depended on three aspects. First, the type of leadership behavior employees should execute: leadership behaviors focused on relations or change were more popular than those externally or task-focused. Also, employees preferred behaviors that take fewer hours and benefit others. Second, willingness depended on the person who was asked: male or younger employees were more willing than female or older employees. Third, willingness depended on the context they were in: employees were more willing to execute shared leadership during a crisis. In this study, we also discuss some of the possible systemic causes that may underlie differences across employees.

Chapter 6: Motivation activation to increase ethical reporting

In a survey experiment (n = 11,728), we assessed the potential of motivation activation as an intervention to increase ethical reporting. A recent study showed that this can work, and we aimed to increase our understanding of the mechanisms of such behavioral change by comparing two types of motivation (public service motivation and prosocial motivation), two types of ethical reporting (reporting colleagues or patients) and analyzing whether effects differed for different levels of motivation. We found that both motivations could increase willingness to report the wrongdoings of patients but not colleagues. The effects that we found were significant but small. Besides, we found that activation of prosocial motivation was more successful for respondents with lower levels of that motivation. These results show that depending on the type and level of motivation and the type of outcome, the effectiveness of behavioral interventions can differ.

Chapter 7: Autonomy-preserving nudges to reduce email use

Across a survey experiment (n = 4,112) and a quasi-field experiment (n = ±1,189), we studied whether nudges—subtle stimuli to change behaviors—could be autonomy-preserving and effective tools to reduce email use among healthcare employees. We developed three nudges: an opinion leader nudge, a rule-of-thumb, and self-nudges. We found that employees deemed the nudges to be autonomy-preserving and effective. Employees also preferred nudges above traditional policy instruments such as a monetary reward. Besides, we found some evidence that actual email use decreased in the healthcare organization where we tested the nudges. This shows that innovative nudges can be both autonomy-preserving and effective.

Conclusions and implications

In this dissertation, we have answered three questions.

First, how can we deepen our understanding of employee well-being in healthcare (***RQ1)***? In **chapter 2,** we made an empirical contribution by differentiating between healthcare employees based on exposure to COVID-19 patients and work and personal characteristics. By documenting the specific impact of a virus, more tailored solutions to increase well-being are possible. In **chapter 3,** we showed how text mining can improve our understanding of the multidimensionality of work engagement. Analyzing employees' self-written stories, our study enabled critical assessment of the work engagement literature.

Second, how can leaders use empowerment to contribute to employee well-being in healthcare (***RQ2)***? Our research identified two boundary conditions: context and employee willingness. In **chapter 4,** we showed that while empowering leadership can increase the well-being of employees in low-intensity crises, this is not the case during a severe crisis. Our findings help nuance the potential of empowerment and explain mixed findings in the literature by stressing how leadership should be adapted to context. In **chapter 5,** we found that employees are generally willing to execute shared leadership. However, willingness depends on the leadership behavior being shared, the person being asked, and the context in which they find themselves. The findings stress the importance of an often-overlooked factor: employee willingness. We also discussed what systemic issues, such as gender inequality, may underlie individual preferences.

Third, how can leaders use behavioral insights to contribute to employee well-being in healthcare (***RQ3)***? We tested two avenues to improve research on behavioral insights: analyzing the mechanisms of behavior change and testing interventions in the field. We did so in the context of well-being-related behaviors. In **chapter 6,** we found that by activating employee motivation, leaders can increase willingness to report patients but not colleagues. Effectiveness also depended on the type and level of motivation. These findings help understand the effectiveness and mechanisms of motivation activation as a behavioral intervention better. In **chapter 7,** we developed innovative nudges and found that they were perceived as autonomy-preserving and effective in reducing email use among healthcare employees. We also found evidence for the effectiveness of the nudges in the field. The findings show that nudges can be a suitable alternative to traditional policy interventions to stimulate employee well-being.

Methodologically, we made two contributions. First, we tested multiple solutions to innovate survey designs. We categorized those solutions as adding temporality (e.g., a longitudinal survey using self-generated identification codes), teaming up with other methods (e.g., combining a survey with administrative data), or adding techniques in the surveys (e.g., using a Bayesian truth serum). Second, we used Open Science practices as much as possible, such as independent ethical review, preregistration and open data.

Finally, there are practical implications: this dissertation can be used to *understand* employee well-being better and *improve* it. That is, we provided multiple concrete tools to improve the measurement of employee well-being (e.g., using text mining) and to use empowerment (e.g., employing shared leadership behaviors) and behavioral insights (e.g., using types of nudges) in a leadership approach.

Hopefully, the insights in this dissertation will contribute to healthier healthcare employment. The well-being of patients and clients in healthcare organizations is unthinkable without appropriate care for the well-being of employees themselves.

Samenvatting in het Nederlands

Introductie

Zorgmedewerkers zijn van onschatbare waarde voor zorgorganisaties. In Nederland draagt hun werk bij aan een relatief goed functionerend gezondheidszorgsysteem en een goede volksgezondheid. Het welzijn van zorgmedewerkers wordt echter steeds meer op de proef gesteld. Een opeenstapeling van werkeisen, zoals lange werktijden en hoge emotionele lasten, gaat gepaard met hogere risico's op burn-out. Als gevolg hiervan hebben zorgmedewerkers een grotere kans op het ontwikkelen van beroepsziekten zoals aandoeningen van het bewegingsapparaat en een hoger ziekteverzuim. Soms overwegen ze zelfs om de zorg te verlaten. Hoe zorgen we voor zorgmedewerkers? In dit proefschrift wilden we ons begrip van het welzijn van zorgmedewerkers verdiepen en strategieën ontwikkelen die het welzijn kunnen helpen verbeteren.

In de literatuur zijn verschillende hulpbronnen geïdentificeerd die een positieve invloed hebben op het welzijn. Eén van de belangrijkste hulpbronnen is leiderschap. Onderzoek toont aan dat leiders het welzijn van medewerkers kunnen verbeteren door hun taakeisen en hulpbronnen te beïnvloeden. Traditioneel wordt leiderschap gedefinieerd als een proces van sociale beïnvloeding richting doelen die mensen in formele leiderschapsrollen nastreven tegenover hun volgers in een organisatie. De afgelopen jaren hebben wetenschappers nieuwe benaderingen van leiderschap ontwikkeld. In dit proefschrift hebben we twee van deze benaderingen bestudeerd: empowerment en gedragsinzichten.

Eerst hebben we empowerment bestudeerd. Volgens sommige wetenschappers zijn medewerkers vaardiger en deskundiger dan ooit tevoren. Daarom pleitten zij voor emancipatie van de beroepsbevolking. Leiders moeten een stap terug doen en medewerkers empoweren. Zo wordt leiderschap een gezamenlijke onderneming van leiders en medewerkers. Het empoweren van medewerkers heeft vaak positieve effecten op het welzijn van zowel leiders als medewerkers. Toch wijzen sommige wetenschappers erop dat het ook een duistere kant kan hebben, bijvoorbeeld door de werkdruk te verhogen. Een concreet voorbeeld van empowerment is dat medewerkers leiderschapsgedrag kunnen vertonen gericht op relaties, door bijvoorbeeld vertrouwenspersoon van het team te worden.

Ten tweede evalueerden we het potentieel van gedragsinzichten. Dit is gebaseerd op het idee dat mensen beperkte besluitvormingsmogelijkheden hebben. Bij onze beslissingen zorgt begrensde rationaliteit ervoor dat we voorspelbare fouten maken. Als reactie daarop hebben wetenschappers theorieën en interventies ontwikkeld op basis van empirisch bewijs van hoe menselijke cognitie werkt. Leiders kunnen deze gedragsinzichten gebruiken om medewerkers te stimuleren beslissingen te nemen die hun welzijn verbeteren. Een voorbeeld is dat leiders kunnen communiceren dat ze minder gaan emailen, wat een nudge is voor werknemers om te volgen.

Onderzoeksvragen

We wilden ons begrip van het welzijn van zorgmedewerkers en hoe leiderschap daaraan kan bijdragen vergroten door onderzoeken te ontwikkelen die een verscheidenheid aan onderwerpen behandelen en een breed scala aan mogelijke benaderingen en hulpmiddelen laten zien. Drie onderzoeksvragen hebben dit proces begeleid.

Onderzoeksvraag 1: *Hoe kunnen we ons inzicht in het welzijn van zorgmedewerkers verdiepen?*

In de hoofdstukken 2 en 3 hebben we twee lacunes in de literatuur over het welzijn van zorgmedewerkers aangepakt door onderscheid te maken tussen groepen zorgmedewerkers en door de meting van het welzijn te innoveren. In **hoofdstuk 2** analyseerden we hoe de ervaringen van Nederlandse zorgmedewerkers varieerden tijdens de COVID-19-crisis. En in **hoofdstuk 3** presenteerden we textmining als een innovatieve methode om het welzijn van zorgmedewerkers te analyseren.

Onderzoeksvraag 2: *Hoe kunnen leiders empowerment gebruiken om bij te dragen aan het welzijn van zorgmedewerkers?*

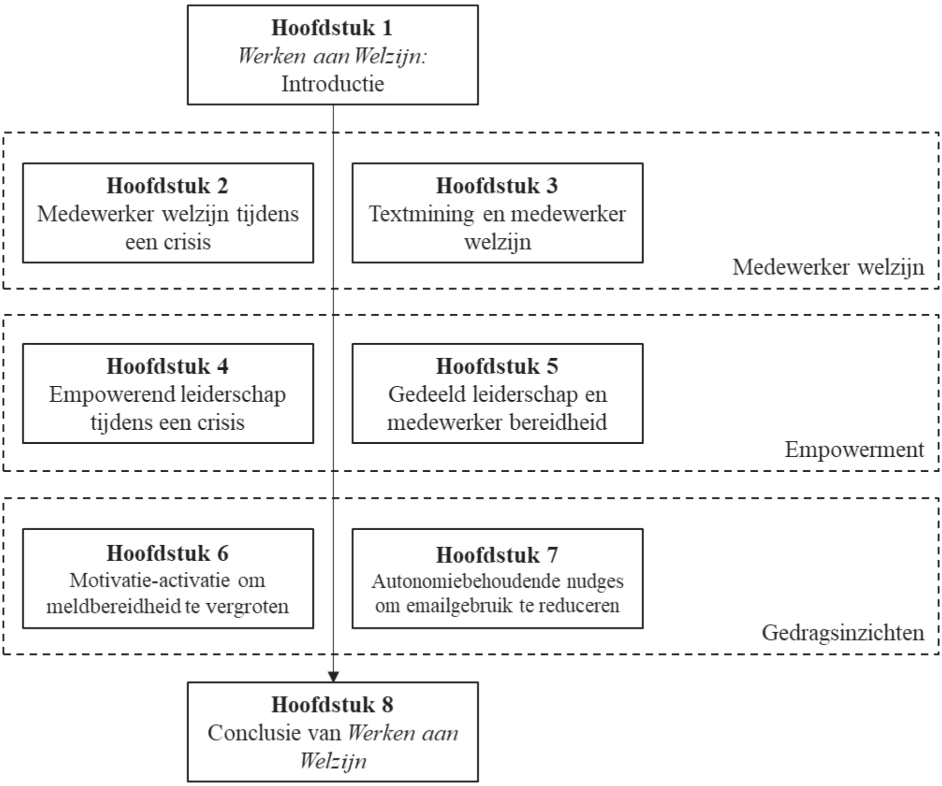
In de hoofdstukken 4 en 5 hebben we twee leiderschapsstijlen bestudeerd die gerelateerd zijn aan empowerment: empowerend leiderschap en gedeeld leiderschap. De literatuur beschrijft vooral positieve effecten van deze stijlen, maar er zijn beperkingen: weinig wetenschappers hebben de rol van context en de bereidheid van medewerkers zelf bestudeerd. In **hoofdstuk 4** onderzochten we of empowerend leiderschap ook een positief effect heeft op het welzijn van medewerkers in een specifieke context: tijdens een crisis. En in **hoofdstuk 5** hebben we onderzocht welke factoren van invloed zijn op de individuele bereidheid van medewerkers om gedeeld leiderschap uit te oefenen.

Onderzoeksvraag 3: *Hoe kunnen leiders gedragsinzichten gebruiken om bij te dragen aan het welzijn van zorgmedewerkers?*

In de hoofdstukken 6 en 7 hebben we het potentieel van gedragsinzichten onderzocht. Om dit te doen hebben wetenschappers gesuggereerd dat we de mechanismen van gedragsverandering moeten bestuderen en interventies rigoureus en in de praktijk moeten testen. We hebben deze lacunes geadresseerd in de context van gedrag gerelateerd aan het welzijn van zorgmedewerkers. In **hoofdstuk 6** hebben we gebruik gemaakt van een recent onderzoek waaruit bleek dat je de motivatie van zorgmedewerkers kunt activeren om daarmee de bereidheid om misstanden te melden te vergroten. We hebben getest of dit afhangt van het type en niveau van de motivatie, en van wie de dader is. In **hoofdstuk 7** hebben we getest of innovatieve nudges ervaren worden als autonomie-behoudend en ze emailgebruik onder zorgmedewerkers kunnen terugdringen.

Figuur 1 geeft een overzicht van dit proefschrift.

Figuur 1 *Overzicht*



Hoofdstuk 2: Medewerker welzijn tijdens een crisis

In een enquête onder 7.208 zorgmedewerkers onderzochten we of zorgmedewerkers die met COVID-19-patiënten werkten een lager welzijn ervoeren dan andere zorgmedewerkers. Ook onderzochten we welke persoonlijke en werkkenmerken de verschillen verklaren binnen de groep zorgmedewerkers die met COVID-19-patiënten werken. We ontdekten dat zorgmedewerkers die in direct contact staan met COVID-19-patiënten meer slaapproblemen en meer fysieke uitputting rapporteerden. Bovendien, binnen die groep zorgmedewerkers rapporteerden zij die vrouw zijn, alleen wonen, geen leiderschapsrol hebben of niet over voldoende beschermingsmiddelen beschikken een lager welzijn. Terwijl de fysieke uitputting hoger was onder oudere zorgmedewerkers, was de mentale uitputting hoger onder jongere zorgmedewerkers. Deze resultaten benadrukken het belang van adequate ondersteuning en pleiten ervoor om rekening te houden met de specifieke omstandigheden van de individuele zorgmedewerker.

Hoofdstuk 3: Textmining en medewerker welzijn

In dit hoofdstuk hebben we textmining geïntroduceerd als een methode om het welzijn van medewerkers te bestuderen. We hebben twee enquêtes (onder 5.591 en 4.470 medewerkers) gebruikt om door middel van textmining de zelfgeschreven verhalen van zorgmedewerkers te classificeren in hoge of lage bevlogenheid. We ontdekten dat psychologische kenmerken—zoals positieve en negatieve emoties—een nauwkeurigheid van 60% hebben bij het classificeren. Door de tekstkenmerken te analyseren die bijdragen aan classificatie, konden we ons begrip van bevlogenheid verdiepen door de literatuur te valideren, uit te breiden of in twijfel te trekken. We hebben de literatuur bijvoorbeeld uitgebreid door aan te tonen dat bevlogenheid een taalkundige component heeft: zeer bevlogen medewerkers gebruikten vaker de eerste persoon meervoud ('wij'). Op basis van de resultaten bespraken we hoe toekomstig onderzoek textmining kan gebruiken om het welzijn van medewerkers beter te begrijpen.

Hoofdstuk 4: Empowerend leiderschap tijdens een crisis

In deze studie combineerden we een longitudinaal onderzoek (n = 468) met administratieve gegevens over de COVID-19-ziekenhuisopnamecijfers in alle provincies in een natuurlijk experiment. We onderzochten of het positieve effect van empowerend leiderschap op het welzijn van zorgmedewerkers aanwezig zou zijn tijdens de crisis. We ontdekten dat voor de motivationele en cognitieve dimensies van welzijn het effect van empowerend leiderschap niet varieert tijdens een crisis. Voor de energetische dimensie (bijvoorbeeld mentale uitputting) is het versterken van leiderschap echter minder effectief in een crisis en kan het zelfs het welzijn van medewerkers schaden.

S

Deze bevindingen laten zien dat empowerend leiderschap negatieve effecten kan hebben in specifieke contexten.

Hoofdstuk 5: Gedeeld leiderschap en medewerker bereidheid

In een experiment onder 6.742 zorgmedewerkers vroegen we hen om leiderschapsgedrag te evalueren dat ze in een gedeelde leiderschapsrol zouden kunnen uitvoeren. We ontdekten dat een aanzienlijk deel van de medewerkers bereid was gedeeld leiderschap uit te voeren. We zagen ook dat de bereidheid afhankelijk was van drie aspecten. Ten eerste het leiderschapsgedrag dat medewerkers zouden moeten uitvoeren: leiderschapsgedrag gericht op relaties of verandering was populairder dan extern of taakgericht gedrag. Bovendien gaven medewerkers de voorkeur aan gedrag dat minder uren in beslag neemt en anderen ten goede komt. Ten tweede hing de bereidheid af van de persoon aan wie de vraag werd gesteld: mannelijke of jongere medewerkers waren meer bereid dan vrouwelijke of oudere medewerkers. Ten derde hing de bereidheid af van de context waarin ze zich bevonden: medewerkers waren meer bereid om gedeeld leiderschap uit te oefenen tijdens een crisis. In deze studie bespreken we ook enkele van de mogelijke systemische oorzaken die ten grondslag kunnen liggen aan verschillen tussen medewerkers.

Hoofdstuk 6: Motivatie-activatie om meldbereidheid te vergroten

In een enquête-experiment (n = 11.728) hebben we het nut van motivatie-activatie beoordeeld als een interventie om de bereidheid tot het melden van misstanden te vergroten. Uit een recent onderzoek is gebleken dat dit kan werken. We wilden ons begrip van de mechanismen van dergelijke gedragsverandering vergroten door twee soorten motivatie (motivatie voor publieke dienstverlening en prosociale motivatie) en twee soorten meldingen (het melden van collega's of patiënten) te vergelijken, en te analyseren of de effecten verschilden voor verschillende niveaus van deze motivatie. We ontdekten dat beide motivaties de bereidheid om misstanden te melden kunnen vergroten bij patiënten, maar niet bij collega's. De effecten die we vonden waren significant, maar klein. Bovendien ontdekten we dat het activeren van prosociale motivatie succesvoller was voor respondenten met een lager niveau van die motivatie. Deze resultaten laten zien dat, afhankelijk van het type en niveau van motivatie en het type uitkomst, de effectiviteit van gedragsinterventies kan verschillen.

Hoofdstuk 7: Autonomiebehoudende nudges om emailgebruik te reduceren

Met een enquête-experiment (n = 4.112) en een quasi-veldexperiment (n = ±1.189) hebben we onderzocht of nudges—subtiele stimuli om gedrag te veranderen—autonomiebehoudende en effectieve hulpmiddelen kunnen zijn om het emailgebruik onder zorgmedewerkers te verminderen. We hebben drie nudges ontwikkeld: een opinieleider nudge, een vuistregel, en zelf-nudges. We ontdekten dat medewerkers de nudges als autonomiebehoudend en effectief beschouwden. Medewerkers gaven ook de voorkeur aan nudges boven traditionele beleidsinstrumenten zoals een financiële beloning. Bovendien hebben we enig bewijs gevonden dat het daadwerkelijke emailgebruik daalde in de zorgorganisatie waar we de nudges hebben getest. Dit toont aan dat innovatieve nudges zowel autonomiebehoudend als effectief kunnen zijn.

Conclusies en implicaties

In dit proefschrift hebben we drie onderzoeksvragen beantwoord.

Ten eerste: hoe kunnen we ons inzicht in het welzijn van zorgmedewerkers verdiepen (***onderzoeksvraag 1)***? In **hoofdstuk 2** hebben we een empirische bijdrage geleverd door onderscheid te maken tussen zorgmedewerkers op basis van blootstelling aan COVID-19-patiënten en werk- en persoonlijke kenmerken. Door de specifieke impact van het virus in kaart te brengen, zijn op maat gemaakte oplossingen om het welzijn te verhogen mogelijk. In **hoofdstuk 3** hebben we laten zien hoe textmining ons begrip van de multidimensionaliteit van bevlogenheid kan verbeteren. Door de zelfgeschreven verhalen van medewerkers te analyseren, maakte ons onderzoek een kritische beoordeling van de literatuur over bevlogenheid mogelijk.

Ten tweede: hoe kunnen leiders empowerment gebruiken om bij te dragen aan het welzijn van zorgmedewerkers (***onderzoeksvraag 2)***? Ons onderzoek identificeerde twee randvoorwaarden: context en bereidheid van medewerkers. In **hoofdstuk 4** hebben we laten zien dat empowerend leiderschap weliswaar het welzijn van medewerkers in laag intensieve crises kan vergroten, maar dat dit tijdens een ernstige crisis niet het geval is. Onze bevindingen helpen het potentieel van empowerment te nuanceren en contrasterende bevindingen in de literatuur te verklaren door te benadrukken dat leiderschap moet worden aangepast aan de context. In **hoofdstuk 5** hebben we ontdekt dat medewerkers over het algemeen bereid zijn om gedeeld leiderschap uit te voeren, maar dat die bereidheid afhangt van het leiderschapsgedrag dat wordt gedeeld, de persoon aan wie het wordt gevraagd en de context waarin zij zich bevinden. De bevindingen benadrukken het belang van een vaak over het hoofd geziene factor: de bereidheid van medewerkers. We bespraken ook welke systemische kwesties, zoals genderongelijkheid, ten grondslag kunnen liggen aan individuele voorkeuren.

Ten derde: hoe kunnen leiders gedragsinzichten gebruiken om bij te dragen aan het welzijn van zorgmedewerkers (***onderzoeksvraag 3)***? We hebben twee manieren getest om het onderzoek naar gedragsinzichten te verbeteren: het analyseren van de mechanismen van gedragsverandering en het rigoureus testen van interventies in de praktijk. We deden dit in de context van gedrag gerelateerd aan welzijn. In **hoofdstuk 6** ontdekten we dat leiders, door de motivatie van medewerkers te activeren, de bereidheid kunnen vergroten om misstanden van patiënten te melden, maar niet die van collega's. De effectiviteit hing ook af van het type en niveau van motivatie. Deze bevindingen helpen de effectiviteit en mechanismen van motivatie-activatie als gedragsinterventie beter te begrijpen. In **hoofdstuk 7** ontwikkelden we innovatieve nudges en vonden we dat deze werden gezien als autonomiebehoudend en effectief in het terugdringen van emailgebruik onder zorgmedewerkers. Ook in de praktijk vonden we bewijs voor de effectiviteit van de nudges. De bevindingen laten zien dat nudges een geschikt alternatief kunnen zijn voor traditionele beleidsinterventies om het welzijn van medewerkers te stimuleren.

Methodologisch hebben we twee bijdragen geleverd. Ten eerste hebben we meerdere oplossingen getest om enquêteontwerpen te innoveren. We hebben deze oplossingen gecategoriseerd als: het toevoegen van temporaliteit (bijvoorbeeld een longitudinale enquête met behulp van zelf gegenereerde identificatiecodes), het combineren met andere methoden (bijvoorbeeld het combineren van een enquête met administratieve gegevens), of het toevoegen van technieken aan de enquête (bijvoorbeeld het gebruik van een Bayesiaans waarheidsserum). Ten tweede hebben we zoveel mogelijk gebruik gemaakt van Open Science-praktijken, zoals onafhankelijke ethische toetsing, preregistratie en open data.

Ten slotte zijn er praktische implicaties: dit proefschrift kan worden gebruikt om het welzijn van medewerkers in organisaties beter te *begrijpen* en te *verbeteren*. We hebben meerdere concrete hulpmiddelen aangereikt om het welzijn van medewerkers beter te meten (bijvoorbeeld door gebruik te maken van textmining) en om empowerment (gedeeld leiderschapsgedrag uitvoeren bijvoorbeeld) en gedragsinzichten (het testen van allerlei typen nudges) te gebruiken in een leiderschapsbenadering.

Hopelijk dragen de inzichten in dit proefschrift bij aan gezonder werk in de gezondheidszorg. Het welzijn van patiënten en cliënten in zorgorganisaties is ondenkbaar zonder passende zorg voor het welzijn van de medewerkers zelf.

References

‘t Hart, P. T., & Tummers, L. (2019). *Understanding public leadership.* London: Red Globe Press.

ABF Research (2021). *Prognosemodel Zorg en Welzijn*. Retrieved from: https://prognosemodelzw.databank.nl/dashboard/dashboard-branches/totaal-zorg-en-welzijn--breed-/

Acemoglu, D., Autor, D. H., & Lyle, D. (2004). Women, war, and wages: The effect of female labor supply on the wage structure at midcentury. *Journal of Political Economy, 112*(3), 497-551.

Adams, J. G., & Walls, R. M. (2020). Supporting the health care workforce during the COVID-19 global epidemic. *JAMA, 323*, 1439-1440.

Adriaenssens, J., De Gucht, V., & Maes, S. (2012). The impact of traumatic events on emergency room nurses: findings from a questionnaire survey. *International Journal of Nursing Studies, 49*, 1411-1422.

Aggarwal, C. C., & Zhai, C. (Eds.). (2012). *Mining text data*. Springer Science & Business Media.

Aguinis, H., Joo, H., & Gottfredson, R. K. (2013). What monetary rewards can and cannot do: how to show employees the money. *Business Horizons, 56*(2), 241-249.

Ahearne, M., Mathieu, J., & Rapp, A. (2005). To empower or not to empower your sales force? An empirical examination of the influence of leadership empowerment behavior on customer satisfaction and performance. *Journal of Applied Psychology, 90*(5), 945-955.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179-211.

Albrecht, S. L., & Andreetta, M. (2011). The influence of empowering leadership, empowerment and engagement on affective commitment and turnover intentions in community health service workers. *Leadership in Health Services, 24*(3), 228-237.

Amundsen, S., & Martinsen, Ø. L. (2014). Empowering leadership: Construct clarification, conceptualization, and validation of a new scale. *The Leadership Quarterly, 25*(3), 487-511.

Amundsen, S., & Martinsen, Ø. L. (2015). Linking empowering leadership to job satisfaction, work effort, and creativity: The role of self-leadership and psychological empowerment. *Journal of Leadership & Organizational Studies, 22*(3), 304-323.

Andersen, L., & Holm Pedersen, L. (2012). Public Service Motivation and Professionalism. *International Journal of Public Administration, 35*(1), 46-57.

Andersen, S. C., & Hjortskov, M. (2016). Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory, 26*(4), 647-662.

Andersson, P., & Almqvist, G. (2022). Carrots, sticks, sermons or nudges? Survey evidence of the Swedish general public’s attitude towards different public policy tools. *Behavioural Public Policy*, 1-26.

Andon, P., Free, C., Jidin, R., Monroe, G. S., & Turner, M. J. (2018). The impact of financial incentives and perceptions of seriousness on whistleblowing intention. *Journal of Business Ethics, 151*(1), 165-178.

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association, 90*(430), 431-442.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist’s companion.* Princeton University Press.

Ansell, C., Sørensen, E., & Torfing, J. (2021). The COVID-19 pandemic as a game changer for public administration and leadership? The need for robust governance responses to turbulent problems. *Public Management Review, 23*(7), 949-960.

Antonakis, J., & Day, D.V. (Eds.). (2017). *The nature of leadership*. Newbury Park: Sage publications.

Antonakis, J., Avolio, B. J., & Sivasubramaniam, N. (2003). Context and leadership: An examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly, 14*(3), 261-295.

Antonakis, J., Dessi, R., Fischer, T., Foss, N., Haslam, S.A., Kvaløy, O., Lonati, S., Muthukrishna, M., & Schöttner, A. (2023). Theory in leadership and management. *The Leadership Quarterly.* DOI: 10.1016/j.leaqua.2023.101736

Armitage, C. J., & Conner, M. (2001). Efficacy of the Theory of Planned Behaviour: A Meta-analytic Review. *British Journal of Social Psychology, 40*(4), 471-499.

Arnold, K. A. (2017). Transformational leadership and employee psychological well-being: A review and directions for future research. *Journal of Occupational Health Psychology, 22*(3), 381-393.

Ashforth, B. E., & Mael, F. (1989). Social Identity Theory and the Organization. *Academy of Management Review, 14*(1), 20-39.

Audenaert, M., George, B., Bauwens, R., Decuypere, A., Descamps, A. M., Muylaert, J., ... & Decramer, A. (2020). Empowering leadership, social support, and job crafting in public organizations: A multilevel study. *Public Personnel Management, 49*(3), 367-392.

Avolio, B. J., Sosik, J. J., Jung, D. I., & Berson, Y. (2003). Leadership models, methods, and applications. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of Psychology: Vol. 12. Industrial and Organizational Psychology* (pp. 277-307). New York: Wiley.

Avolio, B. J., Walumbwa, F. O., & Weber, T. J. (2009). Leadership: Current theories, research, and future directions. *Annual Review of Psychology, 60*, 421-449.

Awan, S., Esteve, M., & van Witteloostuijn, A. (2020). Talking the talk, but not walking the walk: A comparison of self-reported and observed prosocial behaviour. *Public Administration, 98*(4), 995-1010.

Badura, K. L., Grijalva, E., Newman, D. A., Yan, T. T., & Jeon, G. (2018). Gender and leadership emergence: A meta-analysis and explanatory model. *Personnel Psychology, 71*(3), 335-367.

Bagcchi, S. (2020). Stigma during the COVID-19 pandemic. *The Lancet. Infectious Diseases, 20*(7), 782.

Baker, M. (2016). Reproducibility crisis. *Nature, 533*(26), 353-366.

Bakker, A. B. (2022). The social psychology of work engagement: State of the field. *Career Development International, 27*(1), 36-53.

Bakker, A. B., & Costa, P. (2014). Chronic job burnout and daily functioning: A theoretical analysis. *Burnout Research, 1*, 112-119.

Bakker, A. B., & Demerouti, E. (2008). Towards a model of work engagement. *Career Development International, 13*, 209-223.

Bakker, A. B., & Demerouti, E. (2017). Job Demands-Resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology, 22*(3), 273-285.

Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. I. (2014). Burnout and work engagement: The JD-R approach. *Annual Review of Organizational Psychology and Organizational Behavior, 1*(1), 389-411.

Bakker, A. B., Demerouti, E., & Sanz-Vergel, A. I. (2023). Job Demands-Resources Theory: Ten Years Later. *Annual Review of Organizational Psychology and Organizational Behavior, 10,* 25-53.

Bakker, A. B., Demerouti, E., & Schaufeli, W. B. (2005). The crossover of burnout and work engagement among working couples. *Human Relations, 58*(5), 661-689.

Bakker, A. B., Hakanen, J. J., Demerouti, E., & Xanthopoulou, D. (2007). Job resources boost work engagement, particularly when job demands are high. *Journal of Educational Psychology, 99*(2), 274-284.

Bakker, A. B., Rodríguez-Muñoz, A., & Sanz Vergel, A. I. (2016). Modelling job crafting behaviours: Implications for work engagement. *Human Relations, 69*(1), 169-189.

Bakker, A. B., Tims, M., & Derks, D. (2012). Proactive personality and job performance: The role of job crafting and work engagement. *Human Relations, 65*(10), 1359-1378.

Balducci, B., & Marinova, D. (2018). Unstructured data in marketing. *Journal of the Academy of Marketing Science, 46*(4), 557-590.

Balebako, R., Leon, P. G., Almuhimedi, H., Kelley, P. G., Mugan, J., Acquisti, A., & Sadeh-Koniecpol, N. (2011). Nudging users towards privacy on mobile devices. *PINC2011.*

Banerjee, S., & John, P. (2021). Nudge plus: incorporating reflection into behavioral public policy. *Behavioural Public Policy*, 5717(332), 1-16.

Banerjee, S., & John, P. (2023). Nudge Plus: Putting Citizens at the Heart of Behavioural Public Policy. In L. Reich & C. Sunstein (Eds.), Forthcoming in the *Research Handbook on Nudges and Society.* Cheltenham: Edward Elgar Publishing.

Banks, G. C., Field, J. G., Oswald, F. L., O'Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2022). Answers to 18 questions about open science practices. In J. L. Barling & K. D. Mavroveli (Eds.), *Key Topics in Psychological Methods* (pp. 111-124). Springer Nature Switzerland.

Barber, L. K., & Santuzzi, A. M. (2015). Please respond ASAP: workplace telepressure and employee recovery. *Journal of Occupational Health Psychology, 20*(2), 17.

Barley, S. R., Meyerson, D. E., & Grodal, S. (2011). E-mail as a source and symbol of stress. *Organization Science, 22*(4), 887-906.

Barnes, C. M., Lucianetti, L., Bhave, D. P., & Christian, M. S. (2015). "You wouldn't like me when I'm sleepy": Leaders' sleep, daily abusive supervision, and work unit engagement. *Academy of Management Journal, 58*(5), 1419-1437.

Baron, R. A., & Neuman, J. H. (1998). Workplace Aggression - the Iceberg beneath the Tip of Workplace Violence: Evidence on Its Forms, Frequency, and Targets. *Public Administration Quarterly, 21*(4), 446-464.

Barry, D. (1991). Managing the bossless team: Lessons in distributed leadership. *Organisational Dynamics, 21*, 31-47.

Bartunek, J.M., & Spreitzer, G.M. (2006). The interdisciplinary career of a popular construct used in management: Empowerment in the late 20th century. *Journal of Management Inquiry, 15*(3), 255-273.

Bauckham, G., Lambert, R., Atance, C. M., Davidson, P. S., Taler, V., & Renoult, L. (2019). Predicting our own and others' future preferences: The role of social distance. *Quarterly Journal of Experimental Psychology, 72*(3), 634-642.

Bednarczuk, M. (2018). Identity and Vote Overreporting by Bureaucrats: Implications for Public Service Motivation. *The American Review of Public Administration, 48*(2), 148-158.

Belkin, L. Y., Becker, W. J., & Conroy, S. A. (2020). The invisible leash: the impact of organizational expectations for email monitoring after-hours on employee resources, well-being, and turnover intentions. *Group and Organization Management, 45*(5), 709-740.

Belle, N., & Cantarelli, P. (2017). What Causes Unethical Behavior? A Meta-analysis to Set an Agenda for Public Administration Research. *Public Administration Review, 77*(3), 327-339.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software, 3*(30), 774.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics, 119*(1), 249-275.

Bhanot, S. P., & Linos, E. (2020). Behavioral public administration: Past, present, and future. *Public Administration Review, 80*(1), 168-171.

Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research, 19*(84), 1-5. https://jmlr.org/papers/v19/18-416.html

Bjørkelo, B., & Bye, H. H. (2014). On the Appropriateness of Research Design: Intended and Actual Whistleblowing. In A. J. Brown, D. Lewis, R. Moberly, & W. Vandekerckhove (Eds.), *International Handbook on Whistleblowing Research* (pp. 133-153). Cheltenham: Edward Elgar Publishing.

Boin, A., & Hart, P. T. (2003). Public leadership in times of crisis: Mission impossible? *Public Administration Review, 63*(5), 544-553.

Boin, A., 't Hart, P., Stern, E., and Sundelius, B. (2016). *The Politics of Crisis Management: Public Leadership under Pressure.* Cambridge: Cambridge University Press.

Bolino, M. C., & Grant, A. M. (2016). The Bright Side of Being Prosocial at Work, and the Dark Side, Too: A Review and Agenda for Research on Other-oriented Motives, Behavior, and Impact in Organizations. *Academy of Management Annals, 10*(1), 599-670.

Borst, R. T., & Knies, E. (2023). Well-being of public servants under pressure: the roles of job demands and personality traits in the health-impairment process. *Review of Public Personnel Administration, 43*(1), 159-184.

Bouwman, R., & Grimmelikhuijsen, S. (2016). Experimental public administration from 1992 to 2014: A systematic literature review and ways forward. *International Journal of Public Sector Management, 29*(4/5), 329-350.

Boyd, C. M., Bakker, A. B., Pignata, S., Winefield, A. H., Gillespie, N., & Stough, C. (2011). A longitudinal test of the job demands-resources model among Australian university academics. *Applied Psychology, 60*(1), 112-140.

Bozeman, B., & Youtie, J. (2020). Robotic bureaucracy: administrative burden and red tape in university research. *Public Administration Review, 80*(1), 157-162.

Bradt, J. (2022). Comparing the effects of behaviorally informed interventions on flood insurance demand: an experimental analysis of 'boosts' and 'nudges'. *Behavioural Public Policy, 6*(3), 485-515.

Breaugh, J., Ritz, A., & Alfes, K. (2018). Work Motivation and Public Service Motivation: Disentangling Varieties of Motivation and Job Satisfaction. *Public Management Review, 20*(10), 1423-1443.

Breevaart, K., & Bakker, A. B. (2018). Daily job demands and employee work engagement: The role of daily transformational leadership behavior. *Journal of Occupational Health Psychology, 23*(3), 338.

Breevaart, K., Bakker, A. B., Hetland, J., & Hetland, H. (2014). The influence of constructive and destructive leadership behaviors on follower burnout. In *Burnout at work* (pp. 102-121). London: Psychology Press.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Brewer, G. A., & Selden, S. C. (1998). Whistle blowers in the federal civil service: New evidence of the public service ethic. *Journal of Public Administration Research and Theory, 8*(3), 413-440.

Brown, R., Duck, J., & Jimmieson, N. (2014). E-mail in the workplace: The role of stress appraisals and normative response pressure in the relationship between e-mail stressors and employee strain. *International Journal of Stress Management, 21*(4), 325-347.

Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Jonsson, M. L., & Rahali, B. (2018). Can nudges be transparent and yet effective? *Journal of Economic Psychology, 65*, 41-59.

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour, 5*(8), 980-989.

Caillier, J. G. (2015). Transformational Leadership and Whistleblowing Attitudes: Is This Relationship Mediated by Organizational Commitment and Public Service Motivation? *The American Review of Public Administration, 45*(4), 458-475.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200-230.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317-372.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata*. College Station: Stata Press.

Cao, J., Wei, J., Zhu, H., Duan, Y., Geng, W., Hong, X., et al. (2020). A study of basic needs and psychological wellbeing of medical workers in the fever clinic of a tertiary general hospital in Beijing during the COVID-19 outbreak. *Psychotherapy and Psychosomatics, 89*, 252-254.

Caraban, A., Karapanos, E., Gonçalves, D., & Campos, P. (2019). 23 ways to nudge: a review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-15.

Card, D. (1992). Using regional variation in wages to measure the effects of the federal minimum wage. *ILR Review, 46*(1), 22-37.

Carli, L. L. (2020). Women, gender equality and COVID-19. *Gender in Management: An International Journal, 35(*7/8), 647-655.

Carson, J. B., Tesluk, P. E., & Marrone, J. A. (2007). Shared leadership in teams: An investigation of antecedent conditions and performance. *Academy of Management Journal, 50*(5), 1217-1234.

CBS (2020a). StatLine. *Werknemers met een baan in de zorg en welzijn; persoonskenmerken, regio*. Den Haag: Centraal Bureau voor de Statistiek. Retrieved from https://azwstatline.cbs.nl/?dl=470D9#/AZW/nl/dataset/24016NED/table.

CBS (2020b). *ICT'ers werken vaakst vanuit huis tijdens coronacrisis*. Den Haag: Centraal Bureau voor de Statistiek. Retrieved from https://www.cbs.nl/nl-nl/nieuws/2020/33/ict-ers-werken-vaakst-vanuit-huis-tijdens-coronacrisis.

CBS (2022c). *Bevolkingsteller*. Den Haag: Centraal Bureau voor de Statistiek. Retrieved from: https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/bevolkingsteller

Cecchinato, M. E., Bird, J., & Cox, A. L. (2014). Personalised email tools: a solution to email overload? In *CHI'14 Workshop: Personalised Behaviour Change Technologies*. ACM Conference on Human Factors in Computing Systems (CHI).

Chaturvedi, S., Zyphur, M. J., Arvey, R. D., Avolio, B. J., & Larsson, G. (2012). The heritability of emergent leadership: Age and gender as moderating factors. *The Leadership Quarterly, 23*(2), 219-232.

Chen, D., Zhang, Y., Ahmad, A. B., & Liu, B. (2023). How to fuel public employees' change-oriented organizational citizenship behavior: A two-wave moderated mediation study. *Review of Public Personnel Administration, 43*(1), 185-208.

Cheong, M., Spain, S. M., Yammarino, F. J., & Yun, S. (2016). Two faces of empowering leadership: Enabling and burdening. *The Leadership Quarterly, 27*(4), 602-616.

Cheong, M., Yammarino, F. J., Dionne, S. D., Spain, S. M., & Tsai, C. Y. (2019). A review of the effectiveness of empowering leadership. *The Leadership Quarterly, 30*(1), 34-58.

Chiu, C. Y. C., Owens, B. P., & Tesluk, P. E. (2016). Initiating and utilizing shared leadership in teams: The role of leader humility, team proactive personality, and team performance capability. *Journal of Applied Psychology, 101*(12), 1705-1718.

Choi, D. L. (2004). Public Service Motivation and Ethical Conduct. *International Review of Public Administration, 8,* 99-106.

Christensen, R. K., & Wright, B. E. (2018). Public Service Motivation and Ethical Behavior: Evidence from Three Experiments. *Journal of Behavioral Public Administration, 1*(1). DOI: 10.30636/jbpa.11.18.

Christian, M. S., & Slaughter, J. E. (2007, August). Work engagement: A meta-analytic review and directions for research in an emerging area. In *Academy of Management Proceedings, 2007*(1) (pp. 1-6). Briarcliff Manor, NY: Academy of Management.

Christian, M. S., Garza, A. S., & Slaughter, J. E. (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology, 64*, 89-136.

Chuang, C. H., Tseng, P. C., Lin, C. Y., Lin, K. H., & Chen, Y. Y. (2016). Burnout in the intensive care unit professionals: A systematic review. *Medicine, 95*(50), e5629.

Clair, T. S., & Cook, T. D. (2015). Difference-in-differences methods in public finance. *National Tax Journal, 68*(2), 319-338.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohn, A., Fehr, E., & Maréchal, M. A. (2014). Business Culture and Dishonesty in the Banking Industry. *Nature, 516*, 86-89.

Cohn, A., Maréchal, M. A., & Noll, T. (2015). Bad Boys: How Criminal Identity Salience Affects Rule Violation. *The Review of Economic Studies, 82*(4), 1289-1308.

Collinson, D. (2011). Critical leadership studies. *The Sage handbook of leadership*, 181-194. Newbury Park: Sage Publications Ltd.

Collinson, D., Smolović Jones, O., & Grint, K. (2018). 'No more heroes': Critical perspectives on leadership romanticism. *Organization Studies, 39*(11), 1625-1647.

Conger, J. A., & Kanungo, R. N. (1988). The empowerment process: Integrating theory and practice. *Academy of Management Review, 13*(3), 471-482.

Conti, C., Fontanesi, L., Lanzara, R., Rosa, I., Doyle, R. L., & Porcelli, P. (2021). Burnout status of Italian healthcare workers during the first COVID-19 pandemic peak period. *Healthcare, 9*(5), 510.

Costa, P. L., Passos, A. M., & Bakker, A. B. (2014). Team work engagement: A model of emergence. *Journal of Occupational and Organizational Psychology, 87*(2), 414-436.

Cowdrey, C. (2022). New Managers, You Can Create a Workplace That Values Mental Health. *Harvard Business Review*. Retrieved from: https://hbr.org/2022/08/new-managers-you-can-create-a-workplace-that-values-mental-health

Crosby, B. C., & Bryson, J. M. (2018). Why leadership of public leadership research matters: and what to do about it. *Public Management Review, 20*(9), 1265-1286.

Culbertson, S. S., Mills, M. J., & Fullagar, C. J. (2012). Work engagement and work-family facilitation: Making homes happier through positive affective spillover. *Human Relations, 65*(9), 1155-1177.

Currie, G., & Lockett, A. (2011). Distributing leadership in health and social care: concertive, conjoint or collective? *International Journal of Management Reviews, 13*(3), 286-300.

D'Innocenzo, L., Mathieu, J. E., & Kukenberger, M. R. (2016). A meta-analysis of different forms of shared leadership-team performance relations. *Journal of Management, 42*(7), 1964-1991.

Dabbish, L. A., & Kraut, R. E. (2006). Email overload at work: an analysis of factors associated with email strain. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, pp. 431-440.

Daley, C., Gubb, J., Clarke, E., & Bidgood, E. (2013). *Healthcare Systems: The Netherlands*. Civitas Health Unit.

Davidai, S., & Shafir, E. (2020). Are 'nudges' getting a fair shot? Joint versus separate evaluation. *Behavioural Public Policy, 4*(3), 273-291.

Day, S., Christensen, L. M., Dalto, J., & Haug, P. (2007). Identification of Trauma Patients at a Level 1 Trauma Center Utilizing Natural Language Processing. *Journal of Trauma Nursing, 14*(2), 79-83.

De Bruin, J. (2019). *Python Record Linkage Toolkit: A toolkit for record linkage and duplicate detection in Python*. Version v0.14. Zenodo. DOI: 10.5281/zenodo.3559043.

De Dreu, C. K., & Nauta, A. (2009). Self-interest and Other-orientation in Organizational Behavior: Implications for Job Performance, Prosocial Behavior, and Personal Initiative. *Journal of Applied Psychology, 94*(4), 913-926.

De Graaf, G. (2010). A Report on Reporting: Why Peers Report Integrity and Law Violations in Public Organizations. *Public Administration Review, 70*(5), 767-779.

De Ridder, D., Feitsma, J., Van den Hoven, M., Kroese, F., Schillemans, T., Verweij, M., & De Vet, E. (2020). Simple nudges that are not so easy. *Behavioural Public Policy*, 1-19.

De Wind, A., Leijten, F. R., Hoekstra, T., Geuskens, G. A., Burdorf, A., & Van Der Beek, A. J. (2017). "Mental retirement?" Trajectories of work engagement preceding retirement among older workers. *Scandinavian Journal of Work Environment & Health*, *43*(1), 34-41.

Decuypere, A., & Schaufeli, W. (2020). Leadership and work engagement: Exploring explanatory mechanisms. *German Journal of Human Resource Management, 34*(1), 69-95.

Dellavigna, S., & Linos, E. (2022). RCTs to scale: comprehensive evidence from two nudge units. *Econometrica, 90*(1), 81-116.

DellaVigna, S., Pope, D., & Vivalt, E. (2019). Predict science to improve science. *Science, 366*(6464), 428-429.

Demerouti, E., & Bakker, A. B. (2023). Job demands-resources theory in times of crises: New propositions. *Organizational Psychology Review, 13*(3), 209-236.

Demerouti, E., Mostert, K., & Bakker, A. B. (2010). Burnout and work engagement: A thorough investigation of the independency of both constructs. *Journal of Occupational Health Psychology, 15*(3), 209-222.

Dennerlein, T., & Kirkman, B. L. (2022). The hidden dark side of empowering leadership: The moderating role of hindrance stressors in explaining when empowering employees can promote moral disengagement and unethical pro-organizational behavior. *Journal of Applied Psychology*, *107*(12), 2220-2242.

DeRue, D. S., Barnes, C. M., & Morgeson, F. P. (2010). Understanding the motivational contingencies of team leadership. *Small Group Research, 41*(5), 621-651.

Devine, T. M., & Aplin, D. G. (1986). Abuse of Authority: The Office of the Special Counsel and Whistleblower Protection. *Antioch LJ, 4*, 5.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint, arXiv:1810.04805.

Dinh, J. E., Lord, R. G., Gardner, W. L., Meuser, J. D., Liden, R. C., & Hu, J. (2014). Leadership theory and research in the new millennium: Current theoretical trends and changing perspectives. *The Leadership Quarterly, 25*(1), 36-62.

Döös, M., & Wilhelmson, L. (2021). Fifty-five years of managerial shared leadership research: A review of an empirical field. *Leadership, 17*(6), 715-746.

Drescher, M. A., Korsgaard, M. A., Welpe, I. M., Picot, A., & Wigand, R. T. (2014). The dynamics of shared leadership: Building trust and enhancing performance. *Journal of Applied Psychology: An International Review, 99(*5), 771-783.

Einarsen, S., Skogstad, A., Rørvik, E., Lande, Å. B., & Nielsen, M. B. (2018). Climate for conflict management exposure to workplace bullying and work engagement: a moderated mediation analysis. *The International Journal of Human Resource Management, 29*(3), 549-570.

Elliott, C., & Stead, V. (2008). Learning from leading women's experience: Towards a sociological understanding. *Leadership, 4*(2), 159-180.

Fausing, M. S., Joensson, T. S., Lewandowski, J., & Bligh, M. (2015). Antecedents of shared leadership: empowering leadership and interdependence. *Leadership & Organization Development Journal, 36(*3), 271-291.

Fischer, C., Siegel, J., Proeller, I., & Drathschmidt, N. (2022). Resilience through digitalisation: How individual and organisational resources affect public employees working from home during the COVID-19 pandemic. *Public Management Review*, *25*(4), 808-835.

Fishbein, M., & Ajzen, I. (2010). *Predicting and Changing Behavior: The Reasoned Action Approach.* New York: Psychology Press.

Fitzgerald, L., Ferlie, E., McGivern, G., & Buchanan, D. (2013). Distributed leadership patterns and service improvement: Evidence and argument from English healthcare. *The Leadership Quarterly, 24*(1), 227-239.

Fitzsimons, D., James, K. T., & Denyer, D. (2011). Alternative approaches for studying shared and distributed leadership. *International Journal of Management Reviews, 13*(3), 313-328.

Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *The Leadership Quarterly, 21*(6), 1005-1034.

Foley, J. A., Chan, E., van Harskamp, N., & Cipolotti, L. (2020). Comfort always: the importance of providing psychological support to neurology staff, patients, and families during COVID-19. *Frontiers in Psychology, 11.* DOI: 10.3389/fpsyg.2020.573296

Follet, M.P. (1924). *Creative Experience.* New York: Longmans Green.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research, 3*, 1289-1305.

Forman, G. (2004). A pitfall and solution in multi-class feature selection for text classification. In: C.E. Brodley, ed. *Machine learning, proceedings of the twenty-first international conference (ICML 2004)* (pp. 38). Banff, Canada: ACM.

Foulk, T. A., Lanaj, K., Tu, M. H., Erez, A., & Archambeau, L. (2018). Heavy is the head that wears the crown: An actor-centric approach to daily psychological power, abusive leader behavior, and perceived incivility. *Academy of Management Journal, 61*(2), 661-684.

Frank, M. R., Cebrian, M., Pickard, G., & Rahwan, I. (2017). Validating Bayesian truth serum in large-scale online human experiments. *PLOS ONE, 12*(5), e0177385.

Franklin, C. L., & Thompson, K. E. (2005). Response Style and Posttraumatic Stress Disorder (PTSD): A Review. *Journal of Trauma & Dissociation, 6*(3), 105-123.

Freeney, Y., & Fellenz, M. R. (2013). Work engagement job design and the role of the social context at work: Exploring antecedents from a relational perspective. *Human Relations, 66*(11), 1427-1445.

Georganta, K., & Montgomery, A. (2016). Exploring fun as a job resource: The enhancing and protecting role of a key modern workplace factor. *International Journal of Applied Positive Psychology, 1*, 107-131.

Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation.* New York City: WW Norton.

Giurge, L. M., & Bohns, V. K. (2021). You don't need to answer right away! Receivers overestimate how quickly senders expect responses to non-urgent work emails. *Organizational Behavior and Human Decision Processes, 167*, 11-128.

Glickman, M. E., Rao, S. R., & Schultz, M. R. (2014). False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical Epidemiology, 67*(8), 850-857.

Gockel, C., & Werth, L. (2010). Measuring and modeling shared leadership: Traditional approaches and new ideas. *Journal of Personnel Psychology, 9*, 172-180.

Gorgievski, M. J., Peeters, P., Rietzschel, E. F., & Bipp, T. (2016). Betrouwbaarheid en Validiteit van de Nederlandse vertaling van de Work Design Questionnaire. *Gedrag en Organisatie, 29*(3), 273-301.

Graham, R. N. J., & Woodhead, T. (2021). Leadership for continuous improvement in healthcare during the time of COVID-19. *Clinical Radiology, 76*(1), 67-72.

Grant, A. M. (2008a). Does Intrinsic Motivation Fuel the Prosocial Fire? Motivational Synergy in Predicting Persistence, Performance, and Productivity. *The Journal of Applied Psychology, 93*(1), 48-58.

Grant, A. M. (2008b). Employees without a Cause: The Motivational Effects of Prosocial Impact in Public Service. *International Public Management Journal, 11*, 4-66.

Grant, A. M. (2009). Putting Self-interest Out of Business? Contributions and Unanswered Questions from Use-inspired Research on Prosocial Motivation. *Industrial and Organizational Psychology, 2*(1), 94-98.

Grant, A. M., & Berg, J. M. (2011). Prosocial Motivation at Work. In K. S. Cameron & G. M. Spreitzer (Eds.), *The Oxford Handbook of Positive Organizational Scholarship* (Chapter 3). Oxford: Oxford University Press.

Grant, A. M., & Wall, T. D. (2009). The neglected science and art of quasi-experimentation: why-to, when-to, and how-to advice for organizational researchers. *Organizational Research Methods, 12*(4), 653-686.

Grant, A. M., Berg, J. M., & Cable, D. M. (2014). Job titles as identity badges: how self-reflective titles can reduce emotional exhaustion. *Academy of Management Journal, 57*(4), 1201-1225.

Grant, A. M., Christianson, M. K., & Price, R. H. (2007). Happiness, health, or relationships? Managerial practices and employee well-being tradeoffs. *Academy of Management Perspectives, 21*(3), 51-63.

Green, P. E., & Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. *Journal of Marketing, 54*(4), 3-19.

Greenberg, N., Docherty, M., Gnanapragasam, S., & Wessely, S. (2020). Managing mental health challenges faced by healthcare workers during COVID-19 pandemic. *BMJ, 368*, m1211.

Grimmelikhuijsen, S., Jilke, S., Olsen, A. L., & Tummers, L. (2017). Behavioral public administration: Combining insights from public administration and psychology. *Public Administration Review, 77*(1), 45-56.

Groeneveld, S., Tummers, L., Bronkhorst, B., Ashikali, T., & Van Thiel, S. (2015). Quantitative methods in public administration: Their use and development through time. *International Public Management Journal, 18*(1), 61-86.

Gummer, T., Roßmann, J., & Silber, H. (2021). Using instructed response items as attention checks in web surveys: Properties and implementation. *Sociological Methods & Research, 50*(1), 238-264.

Günzel-Jensen, F., Jain, A. K., & Kjeldsen, A. M. (2018). Distributed leadership in health care: The role of formal leadership styles and organizational efficacy. *Leadership, 14*(1), 110-133.

Gupta, A., Sharda, R., & Greve, R. A. (2011). You've got email! Does it really matter to process emails now or later? *Information Systems Frontiers, 13*(5), 637-653.

Hagman, W., Erlansddon, A., Dickert, S., Tinghög, G., & Västfjäll, D. (2022). The effect of paternalistic alternatives on attitudes toward default nudges. *Behavioural Public Policy, 6*(1), 95-118.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences, 112*(8), 2395-2400.

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis, 22*(1), 1-30.

Hakanen, J. J., Ropponen, A., Schaufeli, W. B., & De Witte, H. (2019). Who is engaged at work? A large-scale study in 30 European countries. *Journal of Occupational and Environmental Medicine, 61*(5), 373-381.

Hallsworth, M., & Kirkman, E. (2020). *Behavioral insights.* Boston: MIT Press.

Hallsworth, M., Chadborn, T., Sallis, A., Sanders, M., Berry, D., Greaves, F., ... & Davies, S. C. (2016). Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *The Lancet, 387*(10029), 1743-1752.

Handgraaf, M. J., De Jeude, M. A. V. L., & Appelt, K. C. (2013). Public praise vs. private pay: effects of rewards on energy conservation in the workplace. *Ecological Economics, 86*, 86-92.

Hänig, C., Schierle, M., & Trabold, D. (2010). Comparison of structured vs. unstructured data for industrial quality analysis. *In Proceedings of the world congress on engineering and computer science* (Vol. 1). WCECS 2010, San Francisco, USA.

Hansen, J. A., & Tummers, L. (2020). A Systematic Review of Field Experiments in Public Administration. *Public Administration Review, 80*(6), 921-931.

Hansen, P. G. (2016). The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation, 7*(1), 155-174.

Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the manipulation of choice. *European Journal of Risk Regulation, 4*(1), 3-28.

Hassan, S., & Wright, B. E. (2020). The Behavioral Public Administration Movement: A Critical Reflection. *Public Administration Review, 80*(1), 163-167.

Hausman, D. M., & Welch, B. (2010). Debate: To nudge or not to nudge. *Journal of Political Philosophy, 18*(1), 123-136.

Haverland, M., & Yanow, D. (2012). A hitchhiker's guide to the public administration research universe: surviving conversations on methodologies and methods. *Public Administration Review, 72*(3), 401-408.

Hay, C. (2011). Interpreting interpretivism interpreting interpretations: The new hermeneutics of public administration. *Public Administration, 89*(1), 167-182.

Hayes, A. F. (2012). *PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling* [Whitepaper]. Retrieved from http://www.afhayes.com/public/process2012.pdf.

He, Q. (2013). *Text mining and IRT for psychiatric and psychological assessment*. Enschede: University of Twente.

He, Q., Veldkamp, B. P., & de Vries, T. (2012). Screening for posttraumatic stress disorder using verbal features in self-narratives: A text mining approach. *Psychiatry Research, 198*(3), 441-447.

Heck, W. P. (1992). Police Who Snitch: Deviant Actors in a Secret Society. *Deviant Behavior: An Interdisciplinary Journal, 13*(3), 253-270.

Hertwig, R. (2017). When to consider boosting: some rules for policy-makers. *Behavioural Public Policy, 1*(2), 143-161.

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science, 12*(6), 973-986.

Hertz, S. G., & Krettenauer, T. (2016). Does Moral Identity Effectively Predict Moral Behavior? A Meta-analysis. *Review of General Psychology, 20*(2), 129.

Hiller, N. J., Day, D. V., & Vance, R. J. (2006). Collective enactment of leadership roles and team effectiveness: A field study. *The Leadership Quarterly, 17*(4), 387-397.

Hoch, J. E. (2013). Shared leadership and innovation: The role of vertical leadership and employee integrity. *Journal of Business and Psychology, 28*(2), 159-174.

Hofstede, G., Hofstede, G. J. & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind* (Rev. 3rd ed.). New York: McGraw-Hill.

Hogg, M. A. (2018). *Social Identity Theory*. Redwood City: Stanford University Press.

Hood, C. (1991). A public management for all seasons? *Public Administration, 69*(1), 3-19.

Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling, 10*(3), 435-455.

Horiuchi, Y., Markovich, Z., & Yamamoto, T. (2022). Does conjoint analysis mitigate social desirability bias? *Political Analysis, 30*(4), 535-549.

Houghton, J. D., Pearce, C. L., Manz, C. C., Courtright, S., & Stewart, G. L. (2015). Sharing is caring: Toward a model of proactive caring through shared leadership. *Human Resource Management Review, 25*(3), 313-327.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (Eds.). (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies.* London: Sage publications.

Hudson, R. F., Lane, H. B., & Mercer, C. D. (2005). Writing prompts: The role of various priming conditions on the compositional fluency of developing writers. *Reading and Writing, 18*(6), 473-495.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5-86.

International Labour Office (2013). *Protecting workplace safety and health in difficult economic times - the effect of the financial crisis and economic recession on occupational safety and health*. Geneva: International Labour Office.

Jackson, T., Dawson, R., & Wilson, D. (2003). Reducing the effect of email interruptions on employees. *International Journal of Information Management, 23*(1), 55-65.

Jakobsen, M., & Jensen, R. (2015). Common method bias in public management studies. *International Public Management Journal, 18*(1), 3-30.

James, K. T. (2011). *Leadership in context: Lessons from new leadership theory and current leadership development practice.* London: The King's Fund.

James, O., Jilke, S. R., & Van Ryzin, G. G. (2017). Behavioural and experimental public administration: Emerging contributions and new directions. *Public Administration, 95*(4), 865-873.

Jilke, S., & Tummers, L. (2018). Which clients are deserving of help? A theoretical model and experimental test. *Journal of Public Administration Research and Theory, 28*(2), 226-238.

Jilke, S., & Van Ryzin, G. (2017). Survey experiments for public management research. In O. James, S. Jilke, & G. Van Ryzin (Eds.), *Experiments in public administration research: Challenges and opportunities* (pp. 117-138). Cambridge: Cambridge University Press.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524-532.

John, P., Martin, A., & Mikołajczak, G. (2023). Support for behavioral nudges versus alternative policy instruments and their perceived fairness and efficacy. *Regulation & Governance, 17(*2), 363-371.

John, P., Smith, G., & Stoker, G. (2009). Nudge nudge, think think: two strategies for changing civic behaviour. *The Political Quarterly, 80*(3), 361-370.

Jønsson, T., Unterrainer, C., Jeppesen, H. J., & Jain, A. K. (2016). Measuring distributed leadership agency in a hospital context: development and validation of a new scale. *Journal of Health Organization and Management, 30*(6), 908-926.

Jordan, Z., Lockwood, C., Munn, Z., & Aromataris, E. (2019). The updated Joanna Briggs institute model of evidence-based healthcare. *JBI Evidence Implementation, 17*(1), 58-71.

Jurafsky, D., & Martin, J. H. (2017). *Speech and Language Processing: An Introduction to Natural Language Processing* (3rd ed.). New Jersey, NJ: Pearson Prentice Hall.

Kahn, R. L., Wolfe, D. M., Quinn, R. P., Snoek, J. D., & Rosenthal, R. A. (1964). *Organizational stress: Studies in role conflict and ambiguity*. Hoboken: John Wiley.

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York City: Macmillan.

Kahn-Lang, A., & Lang, K. (2020). The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics, 38*(3), 613-620.

Kane-Frieder, R. E., Hochwarter, W. A., & Ferris, G. R. (2014). Terms of engagement: Political boundaries of work engagement-work outcomes relationships. *Human Relations, 67*(3), 357-382.

Kang, M. M., Park, S., & Sorensen, L. C. (2022). Empowering the frontline: internal and external organizational antecedents of teacher empowerment. *Public Management Review, 24*, 1705-1726.

Kao, A., & Poteet, S. R. (2007). *Natural Language Processing and Text Mining.* Berlin: Springer Science & Business Media.

Karren, R. J., and Barringer, M. W. (2002). A review and analysis of the policy-capturing methodology in organizational research: guidelines for research and practice. *Organizational Research Methods, 5*(4), 337-361.

Kelloway, E. K., & Dimoff, J. K. (2017). Leadership interventions to improve well-being. In R. J. Burke & K. M. Page (Eds.), *Research handbook on work and well-being* (pp. 435-452). Cheltenham: Edward Elgar Publishing.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196-217.

Khan, M. H., Williams, J., Williams, P., & French, E. (2022). Post-heroic heroism: Embedded masculinities in media framing of Australian business leadership. *Leadership, 18*(2), 298-327.

Kim, M., Beehr, T. A., & Prewett, M. S. (2018). Employee responses to empowering leadership: A meta-analysis. *Journal of Leadership & Organizational Studies, 25*(3), 257-276.

Kim, S., & Kim, S. (2016). Social Desirability Bias in Measuring Public Service Motivation. *International Public Management Journal, 19*(3), 293-319.

Kim, S., Vandenabeele, W., Wright, B. E., Andersen, L. B., Cerase, F. P., Christensen, R. K., ... De Vivo, P. (2013). Investigating the structure and meaning of public service motivation across populations: Developing an international instrument and addressing issues of measurement invariance. *Journal of Public Administration Research and Theory, 23*(1), 79-102.

Kim, T., & Holzer, M. (2016). Public employees and performance appraisal: A study of antecedents to employees' perception of the process. *Review of Public Personnel Administration, 36*(1), 31-56.

King, G., & Hermodson, A. (2000). Peer Reporting of Coworker Wrongdoing: A Qualitative Analysis of Observer Attitudes in the Decision to Report versus Not Report Unethical Behavior. *Journal of Applied Communication Research, 28*(4), 309-329.

Kniffin, K. M., Narayanan, J., Anseel, F., Antonakis, J., Ashford, S. P., Bakker, A. B., ... & Vugt, M. V. (2021). COVID-19 and the workplace: Implications, issues, and insights for future research and action. *American Psychologist, 76*(1), 63-77.

Knight, C., Patterson, M., & Dawson, J. (2017). Building work engagement: A systematic review and meta-analysis investigating the effectiveness of work engagement interventions. *Journal of Organizational Behavior, 38*(6), 792-812.

Knight, C., Patterson, M., & Dawson, J. (2019). Work engagement interventions can be effective: a systematic review. *European Journal of Work and Organizational Psychology, 28*(3), 348-372.

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018). Text mining in organizational research. *Organizational Research Methods, 21*(3), 733-765.

Kobayashi, V. B., Mol, S. T., Vrolijk, J., & Kismihók, G. (2021). Text mining in career studies: generating insights from unstructured textual data. In *Handbook of Research Methods in Careers*. Cheltenham: Edward Elgar Publishing.

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering, 160*(1), 3-24.

Krause, G. (1996). *Death of Guilds*. New Haven, CT: Yale University Press.

Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications, caveats, and cautions. *Journal of Second Language Writing, 3*(3), 231-255.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*(2), 201-219.

Krueger, C., & Tian, L. (2004). A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing, 6*(2), 151-157.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*, 1-26.

Kung, F. Y., Kwok, N., & Brown, D. J. (2018). Are attention check questions a threat to scale validity? *Applied Psychology: An International Review, 67*(2), 264-283.

Kuoppala, J., Lamminpää, A., Liira, J., & Vainio, H. (2008). Leadership, job well-being, and health effects: A systematic review and a meta-analysis. *Journal of Occupational and Environmental Medicine, 50*(8), 904-915.

Kushlev, K., & Dunn, E. W. (2015). Checking email less frequently reduces stress. *Computers in Human Behavior, 43*, 220-228.

Kwon, I. (2014). Motivation, Discretion, and Corruption. *Journal of Public Administration Research and Theory, 24*(3), 765-794.

La Bella, A., Fronzetti Colladon, A., Battistoni, E., Castellan, S., & Francucci, M. (2018). Assessing perceived organizational leadership styles through twitter text mining. *Journal of the Association for Information Science and Technology, 69*(1), 21-31.

Lai, X., Li, F., & Leung, K. (2013). A Monte Carlo study of the effects of common method variance on significance testing and parameter bias in hierarchical linear modeling. *Organizational Research Methods, 16*(2), 243-269.

Lapworth, L., James, P., & Wylie, N. (2018). Examining Public Service Motivation in the Voluntary Sector: Implications for Public Management. *Public Management Review, 20*(11), 1663-1682.

Lavena, C. F. (2016). Whistle-Blowing: Individual and Organizational Determinants of the Decision to Report Wrongdoing in the Federal Government. *American Review of Public Administration, 46*(1), 113-136.

Lawton, R., & Parker, D. (2002). Barriers to Incident Reporting in a Healthcare System. *Quality & Safety in Health Care, 11*(1), 15-18.

LCPS (2020). *Archief 2020*. Retrieved from https://lcps.nu/category/archief-2020/

LCPS (2021). *Patientenverplaatsingen*. Retrieved from: https://lcps.nu/patientverplaatsingen/.

Lee, A., Willis, S., & Tian, A. W. (2018). Empowering leadership: A meta-analytic examination of incremental contribution, mediation, and moderation. *Journal of Organizational Behavior, 39*(3), 306-325.

Lee, G., Benoit-Bryan, J., & Johnson, T. P. (2012). Survey research in public administration: Assessing mainstream journals with a total survey error framework. *Public Administration Review, 72*(1), 87-97.

Leeper, T. J., S. B. Hobolt, and J. Tilley. (2020). Measuring subgroup preferences in conjoint experiments. *Political Analysis, 28*(2): 207-221.

Lesener, T., Gusy, B., Jochmann, A., & Wolter, C. (2020). The drivers of work engagement: A meta-analytic review of longitudinal evidence. *Work & Stress, 34*(3), 259-278.

Lewis, D. D. (1998). *Naive (Bayes) at forty: The independence assumption in information retrieval. In European conference on machine learning* (pp. 4-15). Berlin: Springer.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News, 2*(3), 18-22. https://CRAN.R-project.org/doc/Rnews/

Lin, Y., Osman, M., & Ashcroft, R. (2017). Nudge: concept, effectiveness, and ethics. *Basic and Applied Social Psychology, 39*(6), 293-306.

Linos, E. (2018). More than Public Service: A Field Experiment on Job Advertisements and Diversity in the Police. *Journal of Public Administration Research and Theory, 28*, 67-85.

LNAZ (2020). *Opschalingsplan COVID-19*. Retrieved from https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2020/06/30/opschalingsplan-covid-19/opschalingsplan-covid-19.pdf/.

Lorente Prieto, L., Salanova Soria, M., Martínez Martínez, I., & Schaufeli, W. (2008). Extension of the job demands-resources model in the prediction of burnout and engagement among teachers over time. *Psicothema, 20*, 354-360.

Loyens, K. (2013). Why Police Officers and Labour Inspectors (Do Not) Blow the Whistle. *Policing: An International Journal of Police Strategies & Management, 36*, 27-50.

Loyens, K., & Maesschalck, J. (2010). Toward a Theoretical Framework for Ethical Decision Making of Street-level Bureaucracy: Existing Models Reconsidered. *Administration & Society, 42*(1), 66-100.

Luce, D. R., and J. W. Tukey. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology, 1*(1), 1-27.

Lund, C. S., & Andersen, L. B. (2023). Professional development leadership in turbulent times: Public administration symposium: Robust politics and governance in turbulent times. *Public Administration, 101*(1), 124-141.

Maesschalck, J., van der Wal, Z., Huberts, L.W.J.C., et al. (2008). Public Service Motivation and Ethical Conduct. In J. Perry & A. Hondeghem (Eds.), *Motivation in Public Management: The Call of Public Service* (pp. 157-176). Oxford: Oxford University Press.

Maier, M., Bartoš, F., Stanley, T. D., Shanks, D. R., Harris, A. J., & Wagenmakers, E. J. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences, 119*(31), e2200300119.

Mäkikangas, A., Feldt, T., Kinnunen, U., & Mauno, S. (2013). Does personality matter? Research on individual differences in occupational well-being. In *Advances in Positive Organizational Psychology* (Vol. 1, pp. 107-143). Bingley, UK: Emerald.

Mäkikangas, A., Hyvönen, K., & Feldt, T. (2017). The energy and identification continua of burnout and work engagement: Developmental profiles over eight years. *Burnout Research, 5*, 44-54.

Manzano García, G., & Ayala Calvo, J. C. (2021). The threat of COVID-19 and its influence on nursing staff burnout. *Journal of Advanced Nursing, 77*(2), 832-844.

Marchiori, D. R., Adriaanse, M. A., & De Ridder, D. T. (2017). Unresolved questions in nudging research: putting the psychology back in nudging. *Social and Personality Psychology Compass, 11*(1), e12297.

May, D. R., Gilson, R. L., & Harter, L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of Occupational and Organizational Psychology, 77*(1), 11-37.

McDonald, G., Jackson, D., Vickers, M. H., & Wilkes, L. (2016). Surviving workplace adversity: A qualitative study of nurses and midwives and their strategies to increase personal resilience. *Journal of Nursing Management, 24*(1), 123-131.

Mehra, A., Smith, B. R., Dixon, A. L., & Robertson, B. (2006). Distributed leadership in teams: The network of leadership perceptions and team performance. *The Leadership Quarterly, 17*(3), 232-245.

Meijer, A. J. (2011). Networked Coproduction of Public Services in Virtual Communities: From a Government-centric to a Community Approach to Public Service Support. *Public Administration Review, 71*(4), 598-607.

Mertens, S., Herberz, M., Hahnel, U. J., & Brosch, T. (2022). The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains. *Proceedings of the National Academy of Sciences, 119*(1), e2107346118.

Mesmer-Magnus, J. R., & Viswesvaran, C. (2005). Whistleblowing in Organizations: An Examination of Correlates of Whistleblowing Intentions, Actions, and Retaliation. *Journal of Business Ethics, 62*(3), 277-297.

Meyer-Sahling, J. H., Mikkelsen, K. S., & Schuster, C. (2019). The Causal Effect of Public Service Motivation on Ethical Behavior in the Public Sector: Evidence from a Large-scale Survey Experiment. *Journal of Public Administration Research and Theory, 29*(3), 445-459.

Meyer-Sahling, J., Schuster, C., & Mikkelsen, K. (2018). *Civil Service Management in Developing Countries: What Works? Evidence from a Survey of 23,000 Public Servants in Africa, Asia, Latin America, and Eastern Europe.* London: Report prepared for the UK-Department for International Development (DFID).

Mhango, M., Dzobo, M., Chitungo, I., & Dzinamarira, T. (2020). COVID-19 risk factors among health workers: a rapid review. *Safety and Health at Work, 11*, 262-265.

Miceli, M., & Near, J. (1992). *Blowing the Whistle: The Organizational and Legal Implications for Companies and Employees*. Lexington: Lexington Books.

Michaelsen, P., Johansson, L. O., & Hedesström, M. (2021). Experiencing default nudges: autonomy, manipulation, and choice-satisfaction as judged by people themselves. *Behavioural Public Policy,* 1-22.

Miethe, T. D., & Rothschild, J. (1994). Whistleblowing and the Control of Organizational Misconduct. *Sociological Inquiry, 64*, 322-324.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed Representations of Words and Phrases and Their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* (Lake Tahoe, Nevada) (NIPS'13). Red Hook: Curran Associates Inc.

Miller, L. J., & Lu, W. (2019). *These Are the World's Healthiest Nations. Bloomberg.* Retrieved from: https://www.bloomberg.com/news/articles/2019-02-24/spain-tops-italy-as-world-s-healthiest-nation-while-u-s-slips#xj4y7vzkg.

Ministerie van OCW (2021). *Trendrapportage Arbeidsmarkt Leraren po, vo en mbo 2021. Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschap.* Retrieved from: https://www.voion.nl/media/4089/trendrapportage-arbeidsmarkt-leraren-po-vo-en-mbo-2021.pdf

Mohindra, R., Ravaki, R., Suri, V., Bhalla, A., & Singh, S. M. (2020). Issues relevant to mental health promotion in frontline health care providers managing quarantined/isolated COVID19 patients. *Asian Journal of Psychiatry, 51*, 102084.

Molenveld, A., Voorberg, W., Van Buuren, A., & Hagen, L. (2021). A qualitative comparative analysis of collaborative governance structures as applied in urban gardens. *Public Management Review, 23(*11), 1683-1704.

Morgeson, F. P., & Humphrey, S. E. (2006). The Work Design Questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of Applied Psychology, 91*(6), 1321.

Mousa, M., Boyle, J., Skouteris, H., Mullins, A. K., Currie, G., Riach, K., & Teede, H. J. (2021). Advancing women in healthcare leadership: A systematic review and meta-synthesis of multi-sector evidence on organisational interventions. *EClinicalMedicine, 39*, 101084.

Münscher, R., Vetter, M., & Scheuerle, T. (2016). A Review and Taxonomy of Choice Architecture Techniques. *Journal of Behavioral Decision Making, 29*(5), 511-524.

Nagtegaal, R., Tummers, L., Noordegraaf, M., & Bekkers, V. (2019). Nudging healthcare professionals towards evidence-based medicine: a systematic scoping review. *Journal of Behavioral Public Administration, 2*, 2. DOI: 10.30636/jbpa.22.71

Nagtegaal, R., Tummers, L., Noordegraaf, M., & Bekkers, V. (2020). Designing to debias: Measuring and reducing public managers' anchoring bias. *Public Administration Review, 80*(4), 565-576.

Near, J. P., & Miceli, M. P. (1985). Organizational Dissidence: The Case of Whistleblowing. *Journal of Business Ethics, 4*(1), 1-16.

Nederlandse Zorgautoriteit (2017). *Marktscan Acute Zorg*. Retrieved from: https://www.venvn.nl/media/5ibgncdf/acute-zorg.pdf.

Niemand, T., & Mai, R. (2018). Flexible cutoff values for fit indices in the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 46*(6), 1148-1172.

Noordegraaf, M., & Abma, T. (2003). Management by measurement? Public management practices amidst ambiguity. *Public Administration, 81*(4), 853-871.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600-2606.

O'Sullivan, P., & Ngau, O. (2014). Whistleblowing: A Critical Philosophical Analysis of the Component Moral Decisions of the Act and Some New Perspectives on Its Moral Significance. *Business Ethics: A European Review, 23*(4), 401-415.

OECD (2021), *Health at a Glance 2021: OECD Indicators.* Paris: OECD Publishing.

Olsen, A. L., Hjorth, F., Harmon, N., & Barfort, S. (2019). Behavioral Dishonesty in the Public Sector. *Journal of Public Administration Research and Theory, 29*(4), 572-590.

Ong, J. J., Bharatendu, C., Goh, Y., Tang, J. Z., Sooi, K. W., Tan, Y. L., et al. (2020). Headaches associated with personal protective equipment: A cross-sectional study among frontline healthcare workers during COVID-19. *Headache: The Journal of Head and Face Pain, 60*, 864-877.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

OPM (US Office of Personnel Management). (2012). *Federal Employee Viewpoint Survey.* Washington, DC: US Office of Personnel Management.

Osman, M. (2020). *Backfiring, Reactance, Boomerang, Spillovers, and Rebound Effects: Can We Learn Anything from Examples Where Nudges Do the Opposite of What They Intended?* Retrieved from: https://psyarxiv.com/ae756/download?format=pdf

Ospina, S. M. (2017). Collective leadership and context in public administration: Bridging public leadership research and leadership studies. *Public Administration Review, 77*(2), 275-287.

Ouweneel, E., Le Blanc, P. M., Schaufeli, W. B., & van Wijhe, C. I. (2012). Good morning good day: A diary study on positive emotions hope and work engagement. *Human Relations, 65*(9), 1129-1154.

Pak, S., Kramer, A., Lee, Y., & Kim, K. J. (2022). The impact of work hours on work-to-family enrichment and conflict through energy processes: a meta-analysis. *Journal of Organizational Behavior, 43*(4), 709-743.

Paletta, A. (2012). Public governance and school performance: Improving student learning through collaborative public management. *Public Management Review, 14*(8), 1125-1151.

Pang, D., Eichstaedt, J. C., Buffone, A., Slaff, B. R., Ruch, W., & Ungar, L. H. (2020). The language of character strengths: Predicting morally valued traits on social media. *Journal of Personality, 88*(2), 287-306.

Park, J. G., Kim, J. S., Yoon, S. W., & Joo, B. K. (2017). The effects of empowering leadership on psychological well-being and job engagement. *Leadership & Organization Development Journal, 38*(3), 350-367.

Parker, S. K., Wall, T. D., & Cordery, J. L. (2001). Future work design research and practice: Towards an elaborated model of work design. *Journal of Occupational and Organizational Psychology, 74*(4), 413-440.

Patel, R. S., Bachu, R., Adikey, A., Malik, M., & Shah, M. (2018). Factors related to physician burnout and its consequences: a review. *Behavioral Sciences, 8*(11), 98-105.

Pearce, C. L. (2004). The future of leadership: Combining vertical and shared leadership to transform knowledge work. *Academy of Management Perspectives, 18*(1), 47-57.

Pearce, C. L., & Conger, J. A. (2002). *Shared leadership: Reframing the hows and whys of leadership.* Thousand Oaks: Sage Publications.

Pearce, C. L., & Manz, C. C. (2005). The new silver bullets of leadership: The importance of self- and shared leadership in knowledge work. *Organizational Dynamics, 34*, 130-140.

Pearce, C. L., Conger, J. A., & Locke, E. A. (2008). Shared leadership theory. *The Leadership Quarterly, 19*(5), 622-628.

Pearce, C. L., Sims, H. P., Jr., Cox, J. F., Ball, G., Schnell, E., Smith, K. A., & Trevino, L. (2003). Transactors, transformers and beyond: A multi-method development of a theoretical typology of leadership. *Journal of Management Development, 22*(4), 273–307.

Pearce, C. L., Wood, B. G., & Wassenaar, C. L. (2018). The future of leadership in public universities: is shared leadership the answer? *Public Administration Review, 78*(4), 640-644.

Pearman, A., Hughes, M. L., Smith, E. L., & Neupert, S. D. (2020). Mental health challenges of United States healthcare professionals during COVID-19. *Frontiers in Psychology, 11*. DOI: 10.3389/fpsyg.2020.02065

Pedersen, L. H., Andersen, L. B., & Thomsen, N. (2020). Motivated to Act and Take Responsibility - integrating Insights from Community Psychology in PSM Research. *Public Management Review, 22*(7), 999-1023.

Pedersen, M. J. (2015). Activating the Forces of Public Service Motivation: Evidence from a Low-intensity Randomized Survey Experiment. *Public Administration Review, 75*, 734-746.

Pedersen, M. J., & Stritch, J. M. (2018). RNICE Model: Evaluating the contribution of replication studies in public administration and management research. *Public Administration Review, 78*(4), 606-612.

Peng, S., & Li, H. (2019). *Dishonesty in the Name of Noble Cause: Do Public Service Motivation, Prosocial Motivation, and Employment Sector Play a Role?* Presented at the 17th Public Management Research Conference, Chapel Hill, NC.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015.* Austin, TX: University of Texas at Austin. Retrieved from https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing,* pp. 1532-1543. EMNLP 2014, Doha, Qatar.

Perry, J. L., & Hondeghem, A. (2008). *Motivation in Public Management: The Call of Public Service.* Oxford, UK: Oxford University Press.

Perry, J. L., & Vandenabeele, W. (2008). Behavioral Dynamics: Institutions, Identities, and Self-regulation. In J. L. Perry & A. Hondeghem (Eds.), *Motivation in Public Management: The Call of Public Service* (pp. 56-79). Oxford: Oxford University Press.

Perry, J. L., & Vandenabeele, W. (2015). Public Service Motivation Research: Achievements, Challenges, and Future Directions. *Public Administration Review, 75*(5), 692-699.

Perry, P. O. (2017). *Text Corpus Analysis R package version 0.1.02.* Retrieved from http://corpustext.com

Pershing, J. L. (2003). To Snitch or Not to Snitch? Applying the Concept of Neutralization Techniques to the Enforcement of Occupational Misconduct. *Sociological Perspectives, 46*(2), 149-178.

Piatak, J. S., & Holt, S. B. (2019). Prosocial Behaviors: A Matter of Altruism or Public Service Motivation? *Journal of Public Administration Research and Theory, 30*(3), 504-518.

Pillay, S., Reddy, P. S., & Morgan, D. (2017). Institutional Isomorphism and Whistle-blowing Intentions in Public Sector Institutions. *Public Management Review, 19*(4), 423-442.

Pletzer, J. L., Breevaart, K., & Bakker, A. B. (2023). Constructive and destructive leadership in job demands-resources theory: A meta-analytic test of the motivational and health-impairment pathways. *Organizational Psychology Review*, 20413866231197519.

R

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539-569.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method variance in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903.

Porter, L. W., & McLaughlin, G. B. (2006). Leadership and the organizational context: like the weather? *The Leadership Quarterly, 17*(6), 559-576.

Porter, M. F. (2001). *Snowball: A language for stemming algorithms.* Retrieved from http://snowball.tartarus.org/

Post, C., De Smet, H., Uitdewilligen, S., Schreurs, B., & Leysen, J. (2022). Participative or directive leadership behaviors for decision-making in crisis management teams? *Small Group Research, 53*(5), 692-724.

Power, K. (2020). The COVID-19 pandemic has increased the care burden of women and families. *Sustainability: Science, Practice and Policy, 16*(1), 67-73.

Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science, 306*, 462-466.

Raadschelders, J. C. (2011). The future of the study of public administration: Embedding research object and methodology in epistemology and ontology. *Public Administration Review, 71*(6), 916-924.

Raadschelders, J. C., & Lee, K. H. (2011). Trends in the study of public administration: Empirical and qualitative observations from Public Administration Review, 2000-2009. *Public Administration Review, 71*(1), 19-33.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training.* Preprint retrieved from https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf

Ranney, M. L., Griffeth, V., & Jha, A. K. (2020). Critical supply shortages—the need for ventilators and personal protective equipment during the Covid-19 pandemic. *New England Journal of Medicine, 382*(18), e41.

Reijula, S., & Hertwig, R. (2022). Self-nudging and the citizen choice architect. *Behavioural Public Policy, 6*(1), 119-149.

Reinke, K., & Chamorro-Premuzic, T. (2014). When email use gets out of control: Understanding the relationship between personality and email overload and their impact on burnout and work engagement. *Computers in Human Behavior, 36*, 502-509.

Reith, T. P. (2018). Burnout in United States healthcare professionals: a narrative review. *Cureus, 10*(12), e3681.

Rest, J. R. (1986). *Moral Development: Advances in Research and Theory.* New York: Praeger.

Reynolds, S. J., & Ceranic, T. L. (2007). The Effects of Moral Judgment and Moral Identity on Moral Behavior: An Empirical Examination of the Moral Individual. *Journal of Applied Psychology, 92*(6), 1610-1624.

Rich, B. L., Lepine, J. A., & Crawford, E. R. (2010). Job engagement: Antecedents and effects on job performance. *Academy of Management Journal, 53*(3), 617-635.

Ripoll, G. (2019). Disentangling the relationship between public service motivation and ethics: An interdisciplinary approach. *Perspectives on Public Management and Governance, 2*(1), 21-37.

Ripoll, G., & Schott, C. (2020). Does Public Service Motivation Foster Justification of Unethical Behavior? Evidence from Survey Research among Citizens. *International Public Management Journal, 26*(1), 1-22.

Ritz, A., Brewer, G. A., & Neumann, O. (2016). Public Service Motivation: A Systematic Literature Review and Outlook. *Public Administration Review, 76*(3), 414-426.

Ritz, A., Schott, C., Nitzl, C., & Alfes, K. (2020). Public Service Motivation and Prosocial Motivation: Two Sides of the Same Coin? *Public Management Review, 22*(7), 1-25.

RIVM (2022). *Sterfte.* Retrieved from: https://coronadashboard.rijksoverheid.nl/landelijk/sterfte

RIVM. (2020). *COVID-19 cumulatieve aantallen per gemeente.* Retrieved from https://data.rivm.nl/geonetwork/srv/dut/catalog.search#/metadata/1c0fcd57-1102-4620-9cfa-441e93ea5604?tab=contact.

Roodman, D., Nielsen, M. Ø., MacKinnon, J. G., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *The Stata Journal, 19*(1), 4-60.

Rosen, C. C., Simon, L. S., Gajendran, R. S., Johnson, R. E., Lee, H. W., & Lin, S.-H.-J. (2019). Boxed in by your inbox: implications of daily e-mail demands for managers' leadership behaviors. *Journal of Applied Psychology, 104*(1), 19-33.

Rosenberg, S. D., & Tucker, G. J. (1979). Verbal behavior and schizophrenia: The semantic dimension. *Archives of General Psychiatry, 36*(12), 1331-1337.

Rosenthal, U., Charles, M. T., and 't Hart, P. (1989) Introduction, In Rosenthal, U. and Charles, M. T. P. 't Hart (Eds.), *Coping with Crises: The Management of Disasters, Riots and Terrorism* (3-33), Springfield: Charles C. Thomas.

Ryan, M. K., Haslam, S. A., Morgenroth, T., Rink, F., Stoker, J., & Peters, K. (2016). Getting on top of the glass cliff: Reviewing a decade of evidence, explanations, and impact. *The Leadership Quarterly, 27*(3), 446-455.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination Theory: Basic Psychological Needs in Motivation Development and Wellness.* New York: Guilford Publishing.

Saks, A. M., & Gruman, J. A. (2018). Socialization resources theory and newcomers' work engagement: A new pathway to newcomer socialization. *Career Development International, 23*(1), 12-32.

Schaufeli, W. B. and Bakker, A. B. (2004). *UWES. Utrecht Work Engagement Scale. Preliminary Manual.* Version 1.1. Utrecht: Utrecht University.

Schaufeli, W. B., & Bakker, A. B. (2010). Defining and measuring work engagement: Bringing clarity to the concept. In Bakker, A.B., and Leiter, M.P. *Work engagement: A handbook of essential theory and research* (pp. 10-24). Hove: Psychology Press.

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement, 66*(4), 701-716.

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2022). Work Engagement. A critical assessment of the concept and its measurement. In Ruch et al. *Handbook of Positive Psychology Assessment* (pp. 273-295). Gottingen: Hogrefe Publishing.

Schaufeli, W. B., Bakker, A. B., & Van Rhenen, W. (2009). How changes in job demands and resources predict burnout work engagement and sickness absenteeism. *Journal of Organizational Behavior, 30*, 893-917.

Schaufeli, W. B., Leiter, M. P., Maslach, C., & Jackson, S. E. (1996). The MBI-general survey. In C. Maslach, S. E. Jackson, & M. P. Leiter (Eds.), *Maslach Burnout Inventory Manual* (3rd ed., pp. 19-26). Palo Alto, CA: Consulting Psychologists Press.

Schaufeli, W. B., Salanova, M., González-Romá, V., Bakker, A. B. (2002). The measurement of burnout and engagement: a confirmatory factor analytic approach. *Journal of Happiness Studies, 3*(1), 71-92.

Schaufeli, W. B., Shimazu, A., Hakanen, J., Salanova, M., & De Witte, H. (2017). An ultra-short measure for work engagement. *European Journal of Psychological Assessment, 35*(4), 577-591.

Schaufeli, W. B., Taris, T. W., Le Blanc, P., Peeters, M., Bakker, A. B., & De Jonge, J. (2001). Maakt arbeid gezond? Op zoek naar de bevlogen werknemer [Does work make people happy? In search of the engaged worker]. *Psycholoog, 36*, 422-428.

Schmidt, J. E. T., & Groeneveld, S. M. (2021). Setting sail in a storm: leadership in times of cutbacks. *Public Management Review, 23*(1), 112-134.

Schneider, E. C. et al. (2021). *Mirror, Mirror 2021 — Reflecting Poorly: Health Care in the U.S. Compared to Other High-Income Countries.* New York: Commonwealth Fund.

Schnell, R., Bachteler, T., & Reiher, J. (2010). Improving the use of self-generated identification codes. *Evaluation Review, 34*(5), 391-418.

Schoenegger, P. (2023). Experimental philosophy and the incentivisation challenge: a proposed application of the Bayesian Truth Serum. *Review of Philosophy and Psychology, 14*(1), 295-320.

Schott, C. (2015). *Playing a Role-but Which One? How Public Service Motivation and Professionalism Affect Decision-making in Dilemma Situations.* Doctoral dissertation. Den Haag: Leiden University.

Schott, C., & Ritz, A. (2018). The dark sides of public service motivation: A multi-level theoretical framework. *Perspectives on Public Management and Governance, 1*(1), 29-42.

Schott, C., Neumann, O., Baertschi, M., & Ritz, A. (2019). Public Service Motivation, PM and Altruism: Towards Disentanglement and Conceptual Clarity. *International Journal of Public Administration, 42*(14), 1200-1211.

Schuster, C., Weitzman, L., Sass Mikkelsen, K., Meyer-Sahling, J., Bersch, K., Fukuyama, F., ... & Kay, K. (2020). Responding to COVID-19 through surveys of public servants. *Public Administration Review, 80*(5), 792-796.

Schweiger, S., Müller, B., & Güttel, W. H. (2020). Barriers to leadership development: Why is it so difficult to abandon the hero? *Leadership, 16*(4), 411-433.

Scott, S. S., & Henneman, E. (2017). Underreporting of Medical Errors. *MedSurg Nursing, 26*(3), 211-214.

Searle, R. H., & Rice, C. (2020). Making an Impact in Healthcare Contexts: Insights from a Mixed-methods Study of Professional Misconduct. *European Journal of Work and Organizational Psychology, 30*(4), 470-481.

Sfantou, D. F., Laliotis, A., Patelarou, A. E., Sifaki-Pistolla, D., Matalliotakis, M., & Patelarou, E. (2017, October). Importance of leadership style towards quality of care measures in healthcare settings: A systematic review. *Healthcare, 5*(4), 73.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Houghton: Mifflin and Company.

Shand, R., Parker, S., Liddle, J., Spolander, G., Warwick, L., & Ainsworth, S. (2022). After the applause: understanding public management and public service ethos in the fight against Covid-19. *Public Management Review*, *25*(8), 1-23.

Sharma, P. N., & Kirkman, B. L. (2015). Leveraging leaders: A literature review and future lines of inquiry for empowering leadership research. *Group & Organization Management, 40*(2), 193-237.

Shreffler, J., Petrey, J., & Huecker, M. (2020). The impact of COVID-19 on healthcare worker wellness: A scoping review. *Western Journal of Emergency Medicine, 21*(5), 1059-1066.

Shuck, B., Adelson, J. L., & Reio Jr, T. G. (2017). The employee engagement scale: Initial evidence for construct validity and implications for theory and practice. *Human Resource Management, 56*(6), 953-977.

Siemsen, E., Roth, A., & Oliveira, P. (2010). Common method bias in regression models with linear, quadratic, and interaction effects. *Organizational Research Methods, 13*(3), 456-476.

Sieweke, J., & Santoni, S. (2020). Natural experiments in leadership research: An introduction, review, and guidelines. *The Leadership Quarterly, 31*(1), 101338.

Sikström, S., Pålsson Höök, A., & Kjell, O. (2023). Precise language responses versus easy rating scales—Comparing respondents' views with clinicians' belief of the respondent's views. *PLOS ONE, 18*(2), e0267995.

Silge, J., & Robinson, D. (2016). "tidytext: Text Mining and Analysis Using Tidy Data Principles in R". *JOSS, 1*(3), 37.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69*(1), 99-118.

Sims Jr, H. P., Faraj, S., & Yun, S. (2009). When should a leader be directive or empowering? How to develop your own situational theory of leadership. *Business Horizons, 52*(2), 149-158.

Skakon, J., Nielsen, K., Borg, V., & Guzman, J. (2010). Are leaders' well-being, behaviors and style associated with the affective well-being of their employees? A systematic review of three decades of research. *Work & Stress, 24*(2), 107-139.

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review, 36*(2), 381-403.

Smolović Jones, O., Briley, G., & Woodcock, J. (2022). Exposing and re-placing leadership through workers inquiry. *Leadership, 18*(1), 61-80.

Soane, E., Truss, C., Alfes, K., Shantz, A., Rees, C., & Gatenby, M. (2012). Development and application of a new measure of employee engagement: the ISA Engagement Scale. *Human Resource Development International, 15*(5), 529-547.

Söderbacka, T., Nyholm, L., & Fagerström, L. (2020). Workplace interventions that support older employees' health and work ability-a scoping review. *BMC Health Services Research, 20*(1), 1-9.

Somers, M. J., & Casal, J. C. (1994). Organizational Commitment and Whistleblowing: A Test of the Reformer and the Organization Man Hypotheses. *Group & Organization Management, 19*(3), 270-284.

Spoorthy, M. S., Pratapa, S. K., & Mahant, S. (2020). Mental health problems faced by healthcare workers due to the COVID-19 pandemic—A review. *Asian Journal of Psychiatry, 51*, 102-119.

Stazyk, E. C., & Davis, R. S. (2015). Taking the 'High Road': Does Public Service Motivation Alter Ethical Decision Making Processes? *Public Administration, 93*(3), 627-645.

Steen, T. P., & Rutgers, M. R. (2011). The Double-edged Sword: Public Service Motivation, the Oath of Office and the Backlash of an Instrumental Approach. *Public Management Review, 13*(3), 343-361.

Stets, J. E., & Burke, P. J. (2003). A Sociological Approach to Self and Identity. In M. Leary & J. Tangney (Eds.), *Handbook of Self and Identity* (pp. 107-130). New York City: Guilford Press.

Stichting IZZ (2022). *Over IZZ*. Retrieved from: https://izz.nl/over-izz

Stoker, J. I., Garretsen, H., & Soudis, D. (2019). Tightening the leash after a threat: A multi-level event study on leadership behavior following the financial crisis. *The Leadership Quarterly, 30*(2), 199-214.

Strezhnev, A., Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2019). *Conjoint Survey Design Tool: Software Manual.* Retrieved from https://github.com/astrezhnev/conjointsdt.

Sullivan, H., Williams, P., & Jeffares, S. (2012). Leadership for collaboration: Situated agency in practice. *Public Management Review, 14*(1), 41-66.

Sullivan, M., & Karlsson, J. (1998). The Swedish SF-36 Health Survey III. Evaluation of criterion-based validity: results from normative population. *Journal of Clinical Epidemiology, 51*, 1105-1113.

Sumecki, D., Chipulu, M., & Ojiako, U. (2011). Email overload: exploring the moderating role of the perception of email as a 'business critical' tool. *International Journal of Information Management, 31*(5), 407-414.

Sunstein, C. R. (2016). People prefer system 2 nudges (kind of). *Duke Law Journal, 66*, 121-168.

Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, *70*(4), 1159-1202.

Swidorski, C. (1980). Sample Surveys: Help for the "Out-of-House" Evaluator. *Public Administration Review, 40*(1), 67-71.

Szaszi, B., Higney, A., Charlton, A., Gelman, A., Ziano, I., Aczel, B., ... & Tipton, E. (2022). No reason to expect large and consistent effects of nudge interventions. *Proceedings of the National Academy of Sciences, 119*(31), e2200732119.

Sztompka, P. (1999). *Trust: A Sociological Theory.* Cambridge, U.K: Cambridge University Press.

Tan, B. Y., Chew, N. W., Lee, G. K., Jing, M., Goh, Y., Yeo, L. L., et al. (2020). Psychological impact of the COVID-19 pandemic on health care workers in Singapore. *Annals of Internal Medicine, 173*, 317-320.

Tang, G., Chen, Y., van Knippenberg, D., & Yu, B. (2020). Antecedents and consequences of empowering leadership: Leader power distance, leader perception of team capability, and team innovation. *Journal of Organizational Behavior, 41*(6), 551-566.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology, 29*(1), 24-54.

Tavares, G. M., Sobral, F., Goldszmidt, R., & Araújo, F. (2018). Opening the implicit leadership theories' black box: An experimental approach with conjoint analysis. *Frontiers in Psychology, 9*. DOI: 10.3389/fpsyg.2018.00100

Taylor, J. (2018). Internal Whistle-Blowing in the Public Service: A Matter of Trust. *Public Administration Review, 78*(5), 717-726.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness.* New Haven, CT: Yale University Press.

Thaler, R. H., & Sunstein, C. R. (2021). *Nudge: The Final Edition*. London: Penguin Books.

Thau, M., Mikkelsen, M. F., Hjortskov, M., & Pedersen, M. J. (2021). Question Order Bias Revisited: A Split-ballot Experiment on Satisfaction with Public Services among Experienced and Professional Users. *Public Administration, 99*(1), 189-204.

Thomas, K. W., & Velthouse, B. A. (1990). Cognitive elements of empowerment: An "interpretive" model of intrinsic task motivation. *Academy of Management Review, 15*(4), 666-681.

Tian, M., Risku, M., & Collin, K. (2016). A meta-analysis of distributed leadership from 2002 to 2013: Theory development, empirical evidence and future research focus. *Educational Management Administration & Leadership, 44*(1), 146-164.

Tims, M., Bakker, A. B., & Derks, D. (2015). Examining job crafting from an interpersonal perspective: Is employee job crafting related to the well-being of colleagues? *Applied Psychology: An International Review, 64*(4), 727-753.

TNO (2020). *Arbobalans 2020*. Retrieved from: https://wp.monitorarbeid.tno.nl/wp-content/uploads/2021/02/180TNO_Arbobalans2020_V7.pdf

Toegel, G., Kilduff, M., & Anand, N. (2013). Emotion helping by managers: An emergent understanding of discrepant role expectations and outcomes. *Academy of Management Journal, 56(*2), 334-357.

Toppinen-Tanner, S., Ahola, K., Koskinen, A., & Väänänen, A. (2009). Burnout predicts hospitalization for mental and cardiovascular disorders: 10-year prospective results from industrial sector. *Stress and Health: Journal of the International Society for the Investigation of Stress, 25*(4), 287-296.

Torbay, R. (2020). Are we ready for the next pandemic? *Health Affairs, 39*(6), 1104.

Tourish, D. (2015). Some announcements, reaffirming the critical ethos of Leadership, and what we look for in submissions. *Leadership, 11*(2), 135-141.

Tourish, D. (2017). Introduction: Writing differently about leadership. *Leadership, 13*(1), 3-4.

Treviño, L. K., & Victor, B. (1992). Peer Reporting of Unethical Behavior: A Social Context Perspective. *Academy of Management Journal, 35*(1), 38-64.

Treviño, L. T., Weaver, G., & Reynolds, S. J. (2006). Behavioral Ethics in Organizations: A Review. *Journal of Management, 32*(6), 951-990.

Tuckey, M. R., Bakker, A. B., & Dollard, M. F. (2012). Empowering leaders optimize working conditions for engagement: a multilevel study. *Journal of Occupational Health Psychology, 17*(1), 15-27.

Tuckey, M. R., Searle, B., Boyd, C. M., Winefield, A. H., & Winefield, H. R. (2015). Hindrances are not threats: advancing the multidimensionality of work stress. *Journal of Occupational Health Psychology, 20(*2), 131-147.

Tummers, L. G. (2019). Public policy and behavior change. *Public Administration Review, 79*(6), 925-930.

Tummers, L. G. (2023). Nudge in the news: Ethics, effects, and support of nudges. *Public Administration Review, 83*(5), 1015-1036.

Tummers, L. G., & Bakker, A. B. (2021). Leadership and job demands-resources theory: A systematic review. *Frontiers in Psychology, 12*. DOI: 10.3389/fpsyg.2021.722080

Tummers, L. G., Groeneveld, S. M., & Lankhaar, M. (2013). Why do nurses intend to leave their organization? A large-scale analysis in long-term care. *Journal of Advanced Nursing, 69(*12), 2826-2838.

Tummers, L. G., & Knies, E. (2016). Measuring public leadership: Developing scales for four key public leadership roles. *Public Administration, 94*(2), 433-451.

Tummers, L. G., Teo, S., Brunetto, Y., & Teo, S. T. (2016). Workplace Aggression. *International Journal of Public Sector Management, 29*(1), 2-10.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458.

Ulhøi, J. P., & Müller, S. (2014). Mapping the landscape of shared leadership: A review and synthesis. *International Journal of Leadership Studies, 8*(2), 66-87.

United Nations (2022a). *Personal well-being and managing stress for managers and leaders. Factsheet 1: Why is mental health and well-being important?* Retrieved from: https://www.un.org/en/healthy-workforce/files/Fact_Sheet_1.pdf

United Nations (2022b). *Sustainable Development Goals: Goal 3.* Retrieved from: https://sdgs.un.org/goals/goal3

Valente, T. W., & Pumpuang, P. (2007). Identifying opinion leaders to promote behavior change. *Health Education and Behavior, 34*(6), 881-896.

Van Dalen, H. P., Henkens, K., & Schippers, J. (2010). How do employers cope with an ageing workforce? Views from employers and employees. *Demographic Research, 22*, 1015-1036.

Van de Schoot, R., Winter, S. D., Griffioen, E., Grimmelikhuijsen, S., Arts, I., Veen, D., & Tummers, L. G. (2021). The use of questionable research practices to survive in academia examined with expert elicitation, prior-data conflicts, Bayes factors for replication effects, and the Bayes truth serum. *Frontiers in Psychology, 12*. DOI: 10.3389/fpsyg.2021.621547

Van de Voorde, K., Paauwe, J., & Van Veldhoven, M. (2012). Employee well-being and the HRM-organizational performance relationship: A review of quantitative studies. *International Journal of Management Reviews, 14*(4), 391-407.

Van de Walle, S., & Raaphorst, N. (Eds.). (2018). *Inspectors and Enforcement at the Front Line of Government.* Cham: Springer.

Van de Walle, S., & Van Ryzin, G. G. (2011). The Order of Questions in a Survey on Citizen Satisfaction with Public Services: Lessons from a Split-ballot Experiment. *Public Administration, 89*(4), 1436-1450.

Van der Fels, I. (2020). *Resultaten Monitor Gezond werken 2020 tijdens de COVID-19 uitbraak*. Stichting IZZ. Retrieved from https://izz.nl/sites/default/files/downloads/5f6c8a51c0ef7_0.pdf

Van der Fels, I. (2022). *Resultaten Monitor Gezond werken in de zorg 2022*. Stichting IZZ. Retrieved from: https://izz.nl/sites/default/files/downloads/Resultaten%20IZZ%20monitor%20Gezond%20werken%20in%20de%20zorg%202022_1.pdf

Van der Voet, J., Steijn, B., & Kuipers, B. S. (2017). What's in It for Others? The Relationship between Prosocial Motivation and Commitment to Change among Youth Care Professionals. *Public Management Review, 19*(4), 443-462.

Van Roekel, H. (2023). Examining employee willingness to execute shared leadership: The role of leadership behaviour, gender, age, and context. *Leadership, 19*(6), 508-529.

Van Roekel, H., & Schott, C. (2022). Activating employees' motivation to increase intentions to report wrongdoings: evidence from a large-scale survey experiment. *Public Management Review*. DOI: 10.1080/14719037.2021.2015184.

Van Roekel, H., Giurge, L., Schott, C., & Tummers, L. (2023). Nudges can be both autonomy-preserving and effective: Evidence from a survey and quasi-field experiment. *Behavioural Public Policy*. DOI: 10.1017/bpp.2023.18.

Van Roekel, H., Reinhard, J., & Grimmelikhuijsen, S. (2022). Improving hand hygiene in hospitals: Comparing the effect of a nudge and a boost on protocol compliance. *Behavioural Public Policy, 6*(1), 52-74.

Van Roekel, H., van der Fels, I. M., Bakker, A. B., & Tummers, L. G. (2021). Healthcare workers who work with COVID-19 patients are more physically exhausted and have more sleep problems. *Frontiers in Psychology, 11*. DOI: 10.3389/fpsyg.2020.625626

Van Roekel, H., Wigger, E. F. J., Veldkamp, B. P., Bakker, A. B. (2023). What is work engagement? A text mining approach using employees' self-narratives. *Applied Psychology: An International Review.* DOI: 10.1111/apps.12501.

Van Veldhoven, M., Van den Broeck, A., Daniels, K., Bakker, A. B., Tavares, S. M., & Ogbonnaya, C. (2020). Challenging the universality of job resources: Why, when, and for whom are they beneficial? *Applied Psychology: An International Review, 69*(1), 5-29.

Van Wijk, M. (2020). *Arbeidsmarktprofiel van zorg en welzijn.* CBS. Retrieved from: https://www.cbs.nl/nl-nl/longread/statistische-trends/2020/arbeidsmarktprofiel-van-zorg-en-welzijn?onepage=true

Van Wissen, L., & Boot, P. (2017, September). An electronic translation of the LIWC Dictionary into Dutch. In *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference* (pp. 703-715). Portslade: Lexical Computing.

Van Woerkom, M., Bakker, A. B., & Nishii, L. H. (2016). Accumulative job demands and support for strength use: Fine-tuning the job demands-resources model using conservation of resources theory. *Journal of Applied Psychology, 101*(1), 141.

Vandenabeele, W. (2007). Toward a Public Administration Theory of Public Service Motivation: An Institutional Approach. *Public Management Review, 9*(4), 545-556.

Vandenabeele, W., & Jager, S. (2020). Government Calling Revisited: A Survey-Experiment on the Moderating Role of Public Service Motivation in Assessing Employer Attractiveness. *Frontiers in Psychology, 11*. DOI: 10.3389/fpsyg.2020.559011

Vandenabeele, W., & Schott, C. (2020). Public Service Motivation in Public Administrations. In *Oxford Research Encyclopedia of Politics*. Oxford: Oxford University Press.

Vandenabeele, W., & Van de Walle, S. (2008). International differences in public service motivation: Comparing regions across the world. In J. L. Perry & A. Hondeghem (Eds.), *Motivation in Public Management: The Call of Public Service* (pp. 223-244). Oxford: Oxford University Press.

Vandenabeele, W., Ritz, A., & Neumann, O. (2018). Public Service Motivation: State of the Art and Conceptual Cleanup. In E. Ongaro & S. Van Thiel (Eds.), *The Palgrave Handbook of Public Administration and Management in Europe* (pp. 261-278). London: Palgrave Macmillan.

Venema, T. A., Kroese, F. M., & De Ridder, D. T. (2018). I'm still standing: a longitudinal study on the effect of a default nudge. *Psychology & Health, 33*(5), 669-681.

Vitek, V., & Syed Shah, T. (2019). *Implementing a Nudge to Prevent Email Phishing.* Retrieved from: https://www.diva-portal.org/smash/get/diva2:1351517/FULLTEXT01.pdf.

Vogel, D., & Willems, J. (2020). The Effects of Making Public Service Employees Aware of Their Prosocial and Societal Impact: A Microintervention. *Journal of Public Administration Research and Theory, 30*(3), 485-503.

Vogel, D., & Xu, C. (2021). Everything hacked? What is the evidential value of the experimental public administration literature? *Journal of Behavioral Public Administration, 4*(2), 1-17.

Vonk, R. A. A., Hilderink, H. B. M., Plasmans, M. H. D., Kommer, G. J., & Polder, J. J. (2020). *Toekomstverkenning zorguitgaven 2015-2060: Kwantitatief vooronderzoek in opdracht van de Wetenschappelijke Raad voor het Regeringsbeleid (WRR). Deel 1: toekomstprojecties.* Bilthoven: RIVM.

Vugts, A., Van Den Hoven, M., De Vet, E., & Verweij, M. (2020). How autonomy is understood in discussions on the ethics of nudging. *Behavioural Public Policy, 4*(1), 108-123.

Wachner, J., Adriaanse, M. A., & De Ridder, D. T. (2020). And how would that make you feel? How people expect nudges to influence their sense of autonomy. *Frontiers in Psychology, 11*. DOI: 10.3389/fpsyg.2020.607894

Wachner, J., Adriaanse, M. A., & De Ridder, D. T. (2021). The effect of nudges on autonomy in hypothetical and real-life settings. *PLOS ONE, 16*(8), e0256124.

Wang, B., Liu, Y., Qian, J., & Parker, S. K. (2021). Achieving effective remote working during the COVID-19 pandemic: A work design perspective. *Applied Psychology: An International Review, 70*(1), 16-59.

Wang, C. C., Hsieh, H. H., & Wang, Y. D. (2020). Abusive supervision and employee engagement and satisfaction: the mediating role of employee silence. *Personnel Review, 49*(9), 1845-1858.

Wang, J., Zhou, M., & Liu, F. (2020). Reasons for healthcare workers becoming infected with novel coronavirus disease 2019 (COVID-19) in China. *Journal of Hospital Infection, 105*, 100-101.

Wang, N., Zhu, J., Dormann, C., Song, Z., & Bakker, A. B. (2020). The daily motivators: Positive work events, psychological needs satisfaction, and work engagement. *Applied Psychology: An International Review, 69*(2), 508-537.

Warr, P. B. (1987). *Work, unemployment, and mental health*. Oxford: Oxford University Pres

Watts, L. L., & Buckley, M. R. (2017). A Dual-processing Model of Moral Whistleblowing in Organizations. *Journal of Business Ethics, 146*(3), 669-683.

Weaver, R., & Prelec, D. (2013). Creating truth-telling incentives with the Bayesian truth serum. *Journal of Marketing Research, 50*(3), 289-302.

Wei, E., Segall, J., Villanueva, Y., Dang, L. B., Gasca, V. I., Gonzalez, M. P., et al. (2020). Coping with Trauma, celebrating life: reinventing patient and staff support during the COVID-19 pandemic. *Health Affairs, 39*, 1597-1600.

Weintraub, J., Cassell, D., & DePatie, T. P. (2021). Nudging flow through 'SMART'goal setting to decrease stress, increase engagement, and increase performance at work. *Journal of Occupational and Organizational Psychology, 94*(2), 230-258.

Wen, J., Cheng, Y., Hu, X., Yuan, P., Hao, T., & Shi, Y. (2016). Workload, burnout, and medical mistakes among physicians in China: a cross-sectional study. *Bioscience Trends, 10*(1), 27-33.

Wickham, H., François, R., Henry, L., & Müller, K. (2022a). *dplyr: A Grammar of Data Manipulation.* https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr

Wickham, H., Hester, J., & Bryan, J. (2022b). *readr: Read Rectangular Text Data.* https://readr.tidyverse.org, https://github.com/tidyverse/readr

Wickham, H., Miller, E., & Smith, D. (2022c*). haven: Import and Export 'SPSS' 'Stata' and 'SAS' Files.* https://haven.tidyverse.org, https://github.com/tidyverse/haven, https://github.com/WizardMac/ReadStat

Wilkinson, T. M. (2013). Nudging and manipulation. *Political Studies, 61*(2), 341-355.

Wise, L. R. (2000). The Public Service Culture. In R. J. Stillman (Ed.), *Public Administration Concepts and Cases* (pp. 320-329). Boston: Cengage Learning.

Wood, J., Oh, J., Park, J., & Kim, W. (2020). The relationship between work engagement and work-life balance in organizations: A review of the empirical research. *Human Resource Development Review, 19*(3), 240-262.

World Health Organization (2016). *Global strategy on human resources for health: Workforce 2030.* https://apps.who.int/iris/bitstream/handle/10665/250368/9789241511131-eng.pdf

World Health Organization (2019). *Delivered by women, led by men: a gender and equity analysis of the global health and social workforce.* World Health Organization. https://apps.who.int/iris/handle/10665/311322.

Wright, B. E., Christensen, R. K., & Pandey, S. K. (2013). Measuring Public Service Motivation: Exploring the Equivalence of Existing Global Measures. *International Public Management Journal, 16*(2), 197-223.

Wright, B. E., Hassan, S., & Park, J. (2016). Does a Public Service Ethic Encourage Ethical Behavior? Public Service Motivation, Ethical Leadership and the Willingness to Report Ethical Problems. *Public Administration, 94*(3), 647-663.

Wu, Q., Cormican, K., & Chen, G. (2020). A meta-analysis of shared leadership: Antecedents, consequences, and moderators. *Journal of Leadership & Organizational Studies, 27*(1), 49-64.

Xanthopoulou, D., Bakker, A. B., Heuven, E., Demerouti, E., & Schaufeli, W. B. (2008). Working in the sky: a diary study on work engagement among flight attendants. *Journal of Occupational Health Psychology, 13*, 345-356.

Yang, Y., & Hayes, J. A. (2020). Causes and consequences of burnout among mental health professionals: A practice-oriented review of recent empirical literature. *Psychotherapy, 57*(3), 426-436.

Yang, Y., Li, Z., Liang, L., & Zhang, X. (2021). Why and when paradoxical leader behavior impact employee creativity: Thriving at work and psychological safety. *Current Psychology, 40*, 1911-1922.

Yukl, G. (1989). Managerial leadership: A review of theory and research. *Journal of Management, 15*(2), 251-289.

Yukl, G. (2002). *Leadership in organizations* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Yukl, G. (2012). Effective leadership behavior: What we know and what questions need more attention. *Academy of Management Perspectives, 26*(4), 66-85.

Yun, S., Faraj, S., & Sims Jr, H. P. (2005). Should I be directive or empowering? Contingent leadership in high risk and high reliability situations. *Journal of Applied Psychology, 90*(6), 1288-1296.

Zacher, H., Rosing, K., & Frese, M. (2011). Age and leadership: The moderating role of legacy beliefs. *The Leadership Quarterly, 22*(1), 43-50.

Zaller, J., & Feldman, S. (1992). A Simple Theory of the Survey Response: Answering Questions versus Revealing Preferences. *American Journal of Political Science, 36*, 579-616.

Zampetakis, L. A. (2023). Employees' fear at work, job crafting, and work engagement on a daily basis: The case for fear of COVID-19. *Applied Psychology: An International Review, 72*(2), 535-558.

Zhang, F., & Hughes, C. (2020). Beyond p-value: the rigor and power of study. *Global Clinical and Translational Research*, *2*(1), 1-6.

Zhang, Y., Waldman, D. A., Han, Y. L., & Li, X. B. (2015). Paradoxical leader behaviors in people management: Antecedents and consequences. *Academy of Management Journal, 58*(2), 538-566.

Zhou, M., Tang, F., Wang, Y., Nie, H., Zhang, L., You, G., et al. (2020). Knowledge, attitude and practice regarding COVID-19 among health care workers in Henan, China. *Journal of Hospital Infection, 105*, 183-187.

Zhou, P., Yang, X. L., Wang, X. G. et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature, 579*(7798), 270-273.

Zhu, J., Liao, Z., Yam, K. C., & Johnson, R. E. (2018). Shared leadership: A state-of-the-art review and future research agenda. *Journal of Organizational Behavior, 39*(7), 834-852.

Zipparo, L. (1999). Factors Which Deter Public Officials from Reporting Corruption. *Crime, Law, and Social Change, 30*(3), 273-287.

R

List of preregistrations

Note: Chapter 2 presents an exploratory analysis and was therefore not preregistered.

Chapter 3

Van Roekel, H., & Wigger, E. (2022, February 14). Predicting Work Engagement with Employees' Self-Narratives: A Text Mining Approach. https://doi.org/10.17605/OSF.IO/PB8DQ

Chapter 4

Van Roekel, H., & Sieweke, J. (2021, September 21). Employee Wellbeing and Leadership in Healthcare: A Natural Experiment in a Crisis. https://doi.org/10.17605/OSF.IO/8RYN4

Chapter 5

Van Roekel, H. (2023, March 24). Employees' Willingness to Share Leadership: A Role-Focused Model and Conjoint Experiment. https://doi.org/10.17605/OSF.IO/9SYAJ

Chapter 6

Van Roekel, H., Schott, C. (2020, February 3). The Effect of Activating Public Service Motivation and Prosocial Motivation on Ethical Reporting against Colleagues and Patients. OSF preregistration. https://doi.org/10.17605/OSF.IO/SGKM4

Chapter 7

Van Roekel, H., Giurge, L., Schott, C., & Tummers, L. (2022, September 23). Decreasing email overload with autonomy-preserving nudges. Preregistration of the pilot study. https://doi.org/10.17605/OSF.IO/YSGJ2

Van Roekel, H., Giurge, L., Schott, C., & Tummers, L. (2022, September 23). Decreasing email overload with autonomy-preserving nudges. Preregistration of the survey study. https://doi.org/10.17605/OSF.IO/89GBX

Van Roekel, H., Giurge, L., Schott, C., & Tummers, L. (2022, September 23). Decreasing email overload with autonomy-preserving nudges. Preregistration of the field study. https://doi.org/10.17605/OSF.IO/V5YC4

Appendices

Appendix to chapter 2

The dataset can be found here:
https://osf.io/4gk3t/?view_only=0bcc6d0d07d64a9587b2074b32c9475f

Appendices to chapter 3

- **Appendix 1: Included features**

Note: unigrams and bigrams are inductively generated.

LIWC psychological process features

The preselected psychological features are presented in Table 1. More information (like example features, number of words per category and internal consistency) about the features can be found in Pennebaker et al. (2015). Preselection was done based on any resemblance with the definition or items of work engagement, including many features in order to make sure the models assumed little *a priori* structure.

Table 1 *Selected psychological process features (definition from Bakker et al., 2014, p.391)*

Feature	Reflects which aspect of definition/items
Affective processes (positive emotion; negative emotion: anxiety, anger, sadness)	'Positive state of mind'; 'Experiencing a sense of (…) enthusiasm' and 'being (…) happily engrossed', 'mental resilience'; 'I feel happy when I am working intensely'; 'I am enthusiastic about my job'; 'My job inspires me'; 'I am proud on the work that I do'
Social processes	'Being strongly involved in one's work'
Perceptual processes (see, hear, feel)	'Experiencing', 'Being fully concentrated'; 'I feel'
Biological processes (body, health)	'High levels of energy'
Drives (affiliation, achievement, power, reward, risk)	'Being strongly involved in one's work'; 'Experiencing a sense of (…) significance'; 'At my job, I feel strong and vigorous'; 'Experiencing a sense of (…) significance'; 'Experiencing a sense of (…) challenge', 'Even in the face of difficulties'
Time orientations (past focus, present focus, future focus)	Whereby time passes quickly', 'I am immersed in my work', 'I get carried away when I'm working'
Relativity (time)	Whereby time passes quickly', 'I am immersed in my work', 'I get carried away when I'm working'
Personal concerns (work)	'Work-related positive state of mind', e.g.: 'At my work, I feel bursting with energy', 'When I get up in the morning, I feel like going to work'

A

Table 2 presents examples per psychological feature. It also indicates the hierarchy of the features.

Table 2 *Psychological process features (from Pennebaker et al., 2015)*

Feature	Examples
Affective processes	happy, cried
• Positive emotion	love, nice, sweet
• Negative emotion	hurt, ugly, nasty
• Anxiety	worried, fearful
• Anger	hate, kill, annoyed
• Sadness	crying, grief, sad
Social processes	mate, talk, they
Perceptual processes	look, heard, feeling
• See	view, saw, seen
• Hear	listen, hearing
• Feel	feels, touch
Biological processes	eat, blood, pain
• Body	cheek, hands, spit
• Health	clinic, flu, pill
Drives	-
• Affiliation	ally, friend, social
• Achievement	win, success, better
• Power	superior, bully
• Reward	take, prize, benefit
• Risk	danger, doubt
Time orientations	-
• Past focus	ago, did, talked
• Present focus	today, is, now
• Future focus	may, will, soon
Relativity	-
• Time	end, until, season
Personal concerns	-
• Work	job, majors, xerox

LIWC linguistic process features

We selected all features from the linguistic dimensions. Table 3 presents an overview and examples. More information (like number of words per category and internal consistency) about the features can be found in Pennebaker et al. (2015).

Table 3 *Linguistic dimensions (from Pennebaker et al., 2015)*

Feature	Examples
Total function words	it, to, no, very
• Total pronouns	I, them, itself
• Personal pronouns	I, them, her
• 1st pers singular	I, me, mine
• 1st pers plural	we, us, our
• 2nd person	you, your, thou
• 3rd pers singular	she, her, him
• 3rd pers plural	they, their, they'd
• Impersonal pronouns	it, it's, those
• Articles	a, an, the
• Prepositions	to, with, above
• Auxiliary verbs	am, will, have
• Common Adverbs	very, really
• Conjunctions	and, but, whereas
• Negations	no, not, never

Appendix 2: R script

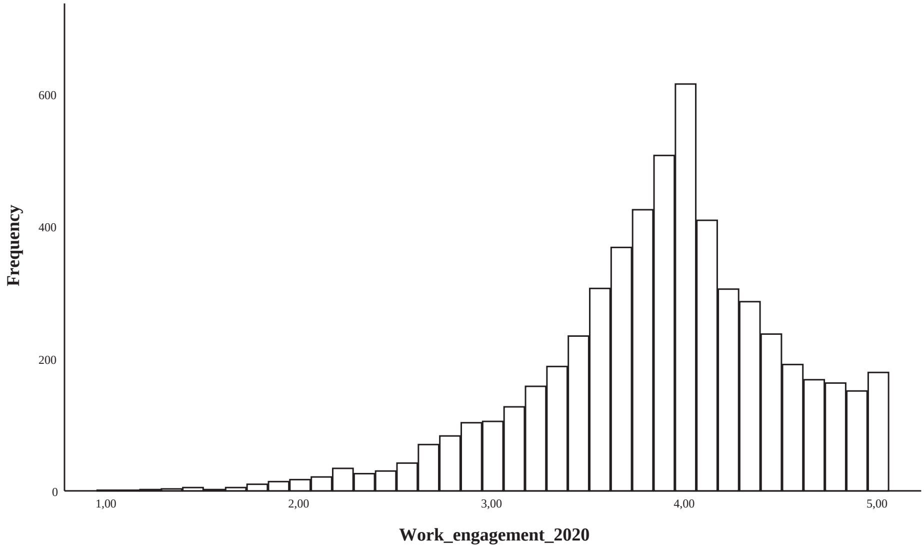
The R script (syntax) can be found here:
https://osf.io/jzdx5/?view_only=acb7ec9693ea43c09c7228b86f00f520

- **Appendix 3: Cutoff**

Exploratory sample 1

Figure 1 shows that the 2020 sample is unbalanced: more healthcare employees tend to be relatively high-engaged. This is in line with earlier findings in the literature on healthcare employees: they are generally high in work engagement (e.g., Hakanen et al., 2019). This causes Figure 1 to be left-skewed (skewness: -.65, kurtosis: .87).

Figure 1 *Scores for work engagement among 2020 sample. n = 5,591; M = 3.87; SD = .63*



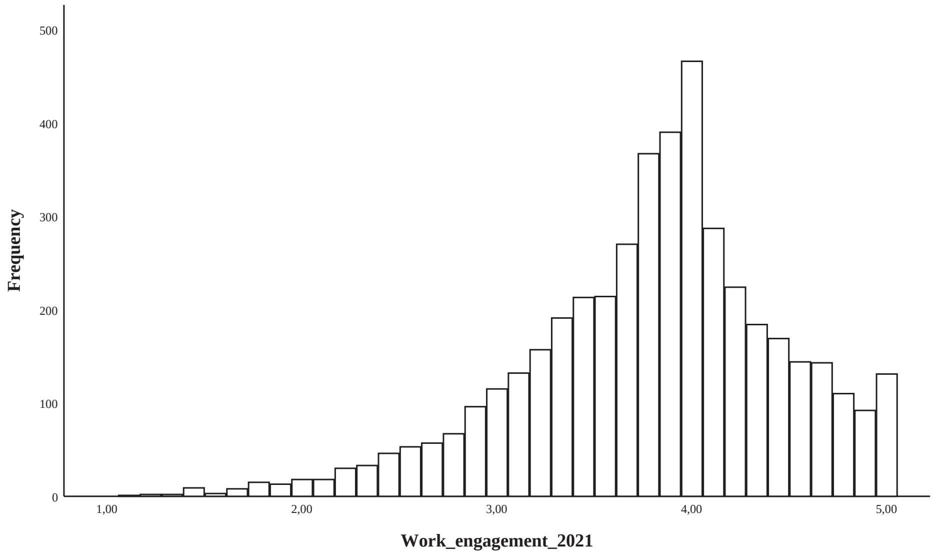
In deciding which employees would be regarded as high-engaged and low-engaged, we considered the effects of this skewness. First, we explored using absolute groups (e.g., scores below 2 and above 4 on the 5-point scale). We found a model using absolute groups did function well due to the skewness in our sample. Due to an overrepresentation of high-engaged employees, the model was insufficiently able to distinguish between low-engaged and high-engaged employees and classified all respondents in the high-engaged group. Resampling in the bottom scoring group in order to artificially inflate the group size did not prove sufficiently fruitful with these large differences in group sizes. Second, we explored relative groups of top scorers and bottom scorers on engagement. This worked well as using a certain percentage of top and bottom scorers guaranteed equal group sizes. We tested four options: 25%, 20%, 10% and 5% of the total sample size. We found that in this exploratory phase a 10% cutoff performed best in distinguishing between high and low scoring respondents. This cutoff was chosen based on the OOB estimates that we derived from this first sample.

Hence, we divided our sample into 10% quantiles and used the lowest and highest quantile as our groups with low and high work engagement.

Confirmatory sample 2

For the confirmatory phase, our second study in which we added a second sample, the same cutoff was set *a priori*. Figure 2 indicates that like sample 1, the work engagement scores in sample 2 are also left-skewed (skewness: -.61, kurtosis: .60).

Figure 2 *Scores for work engagement among 2021 sample. n = 4,470; M = 3.79; SD = .66*



- **Appendix 4: Significant features**

An overview of the significant features can be found here:
https://osf.io/jzdx5/?view_only=acb7ec9693ea43c09c7228b86f00f520

- **Appendix 5: Feature importance**

An overview of the feature importance can be found here:
https://osf.io/jzdx5/?view_only=acb7ec9693ea43c09c7228b86f00f520

- **Appendix 6: Additional analysis**

Variation in work engagement scores in samples

Below we compare work engagement scores in our samples across gender, age and healthcare branch. All sample sizes and work engagement scores across groups can be found in the chapter. Where equal variances could not be assumed, robust tests were reported (e.g., Welch for ANOVA regarding age). The p-levels are set to .05 unless reported otherwise.

Gender

We found that the majority of our sample is female (84.6% in sample 1, 85.4% in sample 2), which is representative of the general population of healthcare employees (84.3%; CBS, 2020a). In sample 1, work engagement is significantly higher among female employees (M = 3.89, SD = .62) compared to male employees (M = 3.78, SD = .67; t (1113.04)= 4.21, p < .001, d = .17). In sample 2, work engagement is also significantly higher among female employees (M = 3.81, SD = .65) compared to male employees (M = 3.70, SD = .72; t (816.28)= 3.37, p < .001, d = .16). Both effect sizes indicate a small effect.

Additionally, we report descriptives regarding gender for the highest and lowest 10% of work engagement scores. For sample 1, within the highest 10% (n = 559), 84.8% is female, and within the lowest 10% (n = 560), 79.3% is female. For sample 2, within the highest 10% (n = 447), 86.4% is female, and within the lowest 10% (n = 447), 79.6% is female.

Age

We found that in our samples most employees are aged between 56 and 65 or 46 and 55 (in sample 1, 8.9% is 35 or younger and 42.7% is older than 55; in sample 2, 6.2% is 35 or younger and 47.4% is older than 55), which is older than the general population of healthcare employees (34% is younger than 35 and 24.2% is older than 55; CBS, 2020a). There is no significant correlation between age and work engagement in either sample (sample 1: r = .02, p = .10; sample 2: r = .02, p = .10). We do find work engagement differs significantly among age groups in sample 1 ($F(5, 295.27) = 3.48$, p = .005, ω^2 = .002) and sample 2 ($F(5, 191) = 5.55$, p < .001, ω^2 = .004). Effect sizes indicate very small effects. Games-Howell post hoc tests indicated that in sample 1, respondents aged 26-35 (M = 3.81, SD = .56) had significantly lower work engagement than respondents aged 46-55 (M = 3.90, SD = .62). In sample 2, respondents aged 26-35 (M = 3.70, SD = .62) also had significantly lower work engagement than respondents aged 46-55 (M = 3.83, SD = .66). Additionally, respondents aged 66 and older (M = 4.16, SD = .56) had significantly higher work engagement than all other groups. However, this result should be interpreted with caution as the group with respondents aged 66 and older included only 36 respondents. In general, the youngest (25 or younger) and oldest (66 or older) groups were small.

Additionally, we report descriptives regarding age for the highest and lowest 10% of work engagement scores. For sample 1, within the highest 10% (n = 559), the mean age is 53.68 (SD = 8.68), and within the lowest 10% (n = 560), the mean age is 52.22 (SD = 10.07). For sample 2, within the highest 10% (n = 447), the mean age is 54.8 (SD = 8.11), and within the lowest 10% (n = 447), the mean age is 53.22 (SD = 9.31).

Healthcare branch

We found that all major healthcare branches are represented in our samples, with a relative overrepresentation of hospitals. We do find work engagement differs significantly among healthcare branches in sample 1 ($F(4, 5586) = 23.43$, $p < .001$, $\eta^2 = .017$) and sample 2 ($F(4, 4465) = 12.89$, $p < .001$, $\eta^2 = .011$). Effect sizes indicate small effects. In sample 1, Bonferroni post hoc tests indicate respondents in mental healthcare (M = 3.76, SD = .62) have significantly lower work engagement than the other healthcare branches. Additionally, respondents in nursing/home care (M = 4.00, SD = .62) have significantly higher work engagement than respondents in the other healthcare branches except the other-category. In sample 2, Bonferroni post hoc tests indicate respondents in nursing/home care (M = 3.91, SD = .66) have significantly higher work engagement than respondents in the other healthcare branches.

Additionally, we report descriptives regarding healthcare branch for the highest and lowest 10% of work engagement scores. For sample 1, within the highest 10% (n = 559), 30.6% works in hospitals, 36.1% works in nursing/home care, 10.2% works in mental healthcare, 15.7% works in disabled care, and 7.3% works in another branch. Within the lowest 10% (n = 560), 37.1% works in hospitals, 17% works in nursing/home care, 21.1% works in mental healthcare, 18.9% works in disabled care, and 5.9% works in another branch.

For sample 2, within the highest 10% (n = 447), 29.3% works in hospitals, 38.3% works in nursing/home care, 11.6% works in mental healthcare, 15% works in disabled care, and 5.8% works in another branch. Within the lowest 10% (n = 447), 41.6% works in hospitals, 20.1% works in nursing/home care, 13.6% works in mental healthcare, 17.7% works in disabled care, and 6.9% works in another branch.

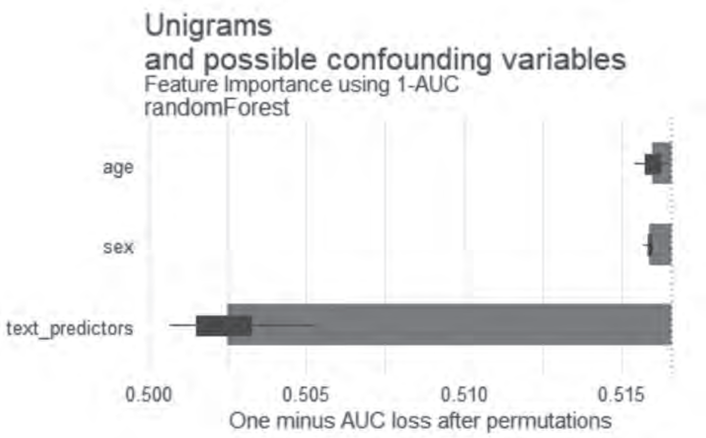
Feature importance: text predictors versus gender and age

Below we use the DALEX package to analyze how age and gender relate to the text-based features when it comes to explaining work engagement (Biecek, 2018). Below analyses have been executed in the context of the confirmatory study on both samples. The analyses were done on a random forest model with age, gender, and text features as input variables and work engagement as output variable. The DALEX package is used to get feature importance scores from the model to compare the importance of each of the features in the model. For this analysis, the text features are grouped together to better understand the importance of age and sex compared to the text features. The results show that gender and age are important features to the models. We discuss implications in the manuscript.

Table 1 *Feature importance unigrams*

Unigrams	Variable	Mean dropout loss
1	full_model	0.5166084
2	text_predictors	0.5025122
3	sex	0.5158909
4	age	0.5159861
5	baseline	0.5016737

Figure 1 *Feature importance unigrams*



A

Table 2 *Feature importance bigrams*

Bigrams	Variable	Mean dropout loss
1	full_model	0.5091970
2	text_predictors	0.5048430
3	age	0.5056989
4	sex	0.5060868
5	baseline	0.4996930

Figure 2 *Feature importance bigrams*

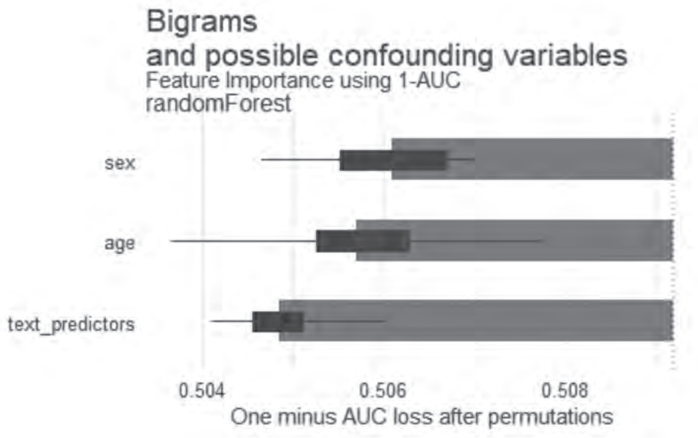


Table 3 *Feature importance psychological features*

Psychological features	Variable	Mean dropout loss
1	full_model	0.5131855
2	text_predictors	0.5021817
3	age	0.5121007
4	sex	0.5125766
5	baseline	0.4995383

Figure 3 *Feature importance psychological features*

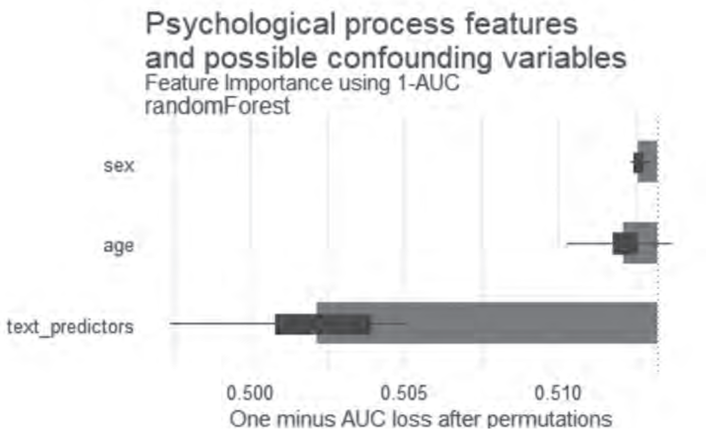
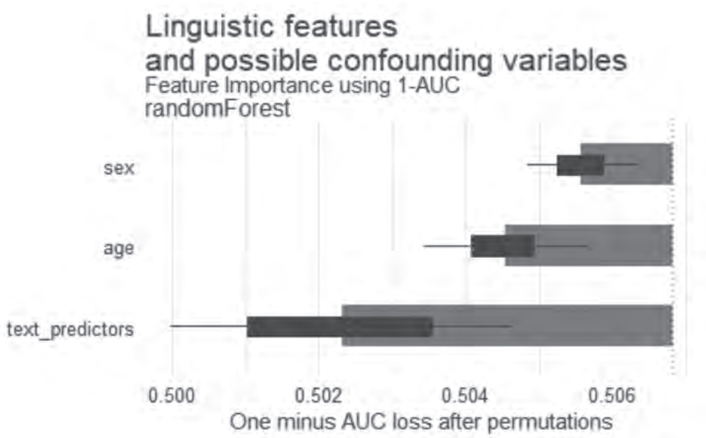


Table 4 *Feature importance linguistic features*

Linguistic features	Variable	Mean dropout loss
1	full_model	0.5068267
2	text_predictors	0.5023220
3	age	0.5045337
4	sex	0.5055635
5	baseline	0.5002473

Figure 4 *Feature importance linguistic features*



A

• Appendix 7: Naive Bayes

Table 1 *Study 2 results executed with Naive Bayes classification model*

		Model 1: unigrams		**Model 2: bigrams**		**Model 3: psychological features**		**Model 4: linguistic features**	
No. of features		156		24		16		6	
Confusion matrix test set (actual/predicted)									
TP	FP	311	184	401	346	303	206	276	232
FN	TN	136	263	46	101	144	241	171	215
Model statistics									
Accuracy		64.21%		56.15%		60.85%		54.92%	
NIR		50.00%		50.00%		50.00%		50.00%	
p-value (Acc > NIR)		< 0.001		< 0.001		< 0.001		0.002	

Note. The confusion matrix presents the numbers for: respondents that score low on work engagement which our model got right (TN), respondents that score low on work engagement which our model got wrong (FP, a type 1 error), respondents that score high on work engagement which our model got wrong (FN, a type 2 error), and respondents that score high on work engagement which our model got right (TP). Note the confusion matrices are reversed compared to the main study, this does not alter the results.

Appendices to chapter 4

- ## Appendix 1: Translations sample items

Table 1 *English versus Dutch sample items of empowering leadership (Ahearne, Mathieu, and Rapp, 2005)*

English	Dutch
My manager makes many decisions together with me.	Mijn leidinggevende maakt veel beslissingen samen met mij.
My manager allows me to do my job my way.	Mijn leidinggevende staat me toe mijn werk op mijn manier te doen.

Note. We used a self-translated version in which a few wordings were changed to adapt to the demographic of the research (e.g., patient needs instead of customer needs).

Table 2 *English versus Dutch sample items of work engagement (UWES-9, Schaufeli et al., 2006; Schaufeli and Bakker, 2004)*

	English	Dutch
Vigor	At my work, I feel bursting with energy.	Op mijn werk bruis ik van energie.
Dedication	I am proud on the work that I do.	Ik ben trots op het werk dat ik doe.
Absorption	I feel happy when I am working intensely.	Wanneer ik heel intensief aan het werk ben, voel ik mij gelukkig.

Table 3 *English versus Dutch sample items of physical and mental exhaustion (adapted from MBI-GS, Schaufeli et al., 1996)*

	English	Dutch
Physical exhaustion	I feel physically exhausted because of my work.	Ik voel me fysiek uitgeput door mijn werk.
Mental exhaustion	I feel mentally exhausted because of my work.	Ik voel me mentaal uitgeput door mijn werk.

- **Appendix 2: Preregistration**

We preregistered this study at the Open Science Framework. The anonymized version is accessible through https://osf.io/8ryn4?view_only=2e1b4a6fe8504964b33419a441142273. The preregistration includes the motive, theoretical model, hypotheses, and methods for our study. The preregistration was registered prior to the analysis of the data. At registration, the preregistration was time-stamped and could not be altered anymore.

As is common practice, we indicate below how the final analysis differs from the preregistration. First, whereas in our preregistration we built up our argument with three analyses, in our final article we decided to focus on one analysis only: the effects of empowering leadership (preregistered ***H3***). The analyses for the direct effects (preregistered ***H1***, Table 1) and for the two-way interactions of the crisis and crisis intensity on well-being (preregistered ***H2***, Table 2) are presented below. For the first analysis, we find only small direct effects of the crisis on well-being. Second, we find only one marginally significant effect for the two-way interactions of the crisis and crisis intensity on well-being.

Second, regarding preregistered ***H3***, we reframed the theoretical model to better express that a) our study focuses on the effect of empowering leadership on well-being and b) the moderating effect a crisis has on this relationship. Hence, empowering leadership was put as independent variable and crisis as moderator. Likewise, we merged hypotheses h3a-e about the dependent variables that measured well-being in the cases they could be merged, because although they were preregistered as separate hypotheses, they referred to the same theoretical expectations. Importantly, these changes did not change the analyses that we conducted. The only thing that changed was the quantity and specific formulation of the hypotheses, as we felt the preregistered hypotheses were insufficiently clear. The description of our preregistration indicates that our final paper aligns with our preregistration: '*many scholars have studied leadership as a vital working condition to improve or sustain occupational wellbeing (e.g., Dimoff and Kelloway, 2017). Healthcare managers may, in times of crisis, affect the extent to which a crisis affects employee wellbeing (e.g., Skakon et al., 2010). Specifically, employees may perceive their manager to empower them in their work through the enactment of their leadership (e.g., Ahearne, Mathieu, and Rapp, 2005), which may act as a buffer to the negative effects of the crisis on wellbeing (Bakker et al., 2014) or strengthen positive effects. Hence, we study how leadership affects the relationship between a crisis and wellbeing.*'

Finally, and we already suggested this in the preregistration, we introduced one more criterium for inclusion: healthcare workers should have been in contact with COVID-19 patients. In Appendix 3 we present an additional analysis for healthcare workers who had not been in contact with COVID-19 patients.

Table 1 *Fixed-effects (within-person) regressions for the effects of the crisis on employee well-being*

	Model 1 Dedication	**Model 2 Absorption**	**Model 3 Vigor**	**Model 4 Physical exhaustion**	**Model 5 Mental exhaustion**
Crisis	-.08*** (.02)	-.04 (.03)	.03 (.02)	-.14*** (.02)	.12*** (.02)
R^2 (within-person)	.02	.003	.003	.05	.03
F	21.79***	1.73	2.61	45.40***	31.14***

Note. $n = 936$; unstandardized coefficients are shown (with clustered standard errors in parentheses); * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

Table 2 *Fixed-effects (within-person) regressions for the two-way interaction effects of the crisis and crisis intensity on well-being*

	Model 1 Dedication	**Model 2 Absorption**	**Model 3 Vigor**	**Model 4 Physical exhaustion**	**Model 5 Mental exhaustion**
Crisis	-0.15** (0.04)	-0.01 (0.07)	0.08 (0.06)	-0.10 (0.06)	0.14* (0.05)
Crisis*crisis intensity	0.0004[a] (0.0002)	-0.0001 (0.0003)	-0.0002 (0.0002)	-0.0002 (0.0002)	-0.0001 (0.0002)
R^2 (within-person)	0.02	0.003	0.004	0.05	0.03
F	26.94***	1.54	1.40	90.51***	30.29***

Note . $n = 936$; unstandardized coefficients are shown (with clustered standard errors in parentheses); the direct effect of crisis intensity is not shown, because it is collinear with the person fixed-effects. [a] $p \leq 0.10$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

- **Appendix 3: Results confirmatory factor analysis and additional analyses**

Confirmation of measurement model

We conducted confirmatory factor analysis (CFA) in MPlus 8.3 to verify our hypothesized factor structure. All items in our measurement model loaded on the respective factor and all factors were free to correlate with each other. Given that the items were skewed, we estimated our model using maximum likelihood estimation with robust standard errors (MLR estimator in MPlus 8.3), which provides standard errors and χ^2 tests that are robust to violations of normality. We tested the data for 2019 and 2020 separately in a CFA and in combination in a multilevel CFA in which observations from 2019 and 2020 are nested within individuals. Our findings (see Table 1) show that our hypothesized six-factor model (empowering leadership, vigor, dedication, absorption, physical exhaustion and mental exhaustion) fitted the data reasonably well (χ^2 = 3013.97; p < .001; df = 838; comparative fit index (CFI) = .87; Tucker-Lewis index (TLI) = .86; standardized root mean square residual: $SRMR_{within}$ = .11; $SRMR_{between}$ = .15; root mean square error of approximation (RMSEA) = .05), especially if we consider that fit indices such as CFI and TLI are downward biased due to the non-normality of our data (Niemand, 2018). All items loaded significantly on the intended factor (p < .05).

We also tested the fit of the six-factor model against the fit of four alternative models: a four-factor model (vigor, dedication and absorption are combined to load on one factor); a three-factor model (physical and mental exhaustion additionally load on one factor); a two-factor model (all dependent variables load on one factor); and a one factor model (all items load on one factor). As shown in Table 1, a χ^2 difference test indicates that the hypothesized six-factor model provides a better fit to the data than any of the alternative models (p < .001).

Table 1 *Results of the Confirmatory Factor Analyses*

Data from 2019										
Model	χ^2	*df*	RMSEA	CFI	TLI	SRMR		SCF	$\Delta\chi^2$	*p*
6 factor model	1637.9	419	0.077	0.855	0.839	0.055		1.120		
4 factor model	1787.8	428	0.081	0.838	0.825	0.063		1.121	147.11	<.001
3 factor model	1949.2	431	0.085	0.82	0.805	0.065		1.122	296.61	<.001
2 factor model	3246.6	433	0.115	0.666	0.641	0.1		1.099	3614.8	<.001
1 factor model	5480.9	434	0.154	0.401	0.358	0.17		1.098	8602.0	<.001
Data from 2020										
Model	χ^2	*df*	RMSEA	CFI	TLI	SRMR		SCF	$\Delta\chi^2$	*p*
6 factor model	1652.4	419	0.077	0.862	0.847	0.054		1.155		
4 factor model	1756.0	428	0.079	0.852	0.839	0.06		1.156	99.72	<.001
3 factor model	1978.3	431	0.085	0.827	0.813	0.062		1.157	310.76	<.001
2 factor model	3174.7	433	0.113	0.693	0.671	0.096		1.147	1878.9	<.001
1 factor model	5640.9	434	0.156	0.418	0.376	0.16		1.122	21146	<.001
Data from 2019 and 2020										
						SRMR				
Model	χ^2	*df*	RMSEA	CFI	TLI	within	between	SCF	$\Delta\chi^2$	*p*
6 factor model	3013.9	838	0.051	0.869	0.855	0.108	0.15	1.000		
4 factor model	3275.7	856	0.054	0.854	0.842	0.139	0.192	0.992	389.98	<.001
3 factor model	3562.7	862	0.056	0.838	0.825	0.19	0.245	1.001	537.74	<.001
2 factor model	4430.9	866	0.065	0.786	0.77	0.283	0.475	0.986	2364.3	<.001
1 factor model	7498.9	868	0.088	0.601	0.573	0.195	0.301	1.021	2902.6	<.001

A

Additional analyses

We conducted several additional analyses to check the robustness of our findings and to analyze potential biases. First, despite the advantages of our DID design—most importantly the exogenous nature of the treatment (i.e., the crisis)—there is nonetheless an endogeneity threat: within-person changes in the level of empowering leadership before and during the crisis may be correlated with crisis intensity. That is, leaders in provinces that were hit hard by the COVID-19 pandemic may purposefully decide to increase (or reduce) the level of empowering leadership, e.g., to improve employee well-being that is negatively affected by the higher crisis intensity. To check this possibility, we regressed within-person changes in empowering leadership between 2019 and 2020 on crisis intensity. Our analysis shows that crisis intensity is unrelated to within-person changes in empowering leadership ($b = 0.0001$, $p = .766$). This indicates that the within-person changes in empowering leadership were unaffected by the intensity of the COVID-19 pandemic within a province. Thus, we conclude that this endogeneity threat is unlikely to bias our estimates.

Second, our findings may be affected by a non-respondent bias. That is, employees with very low levels of employee well-being may be more likely to skip the second wave of our data collection in May and June 2020. To check for this possible bias, we regressed our dependent variables on a time-trend variable that indicates the date when respondent participated in the second wave of our survey. The idea behind this approach is that late respondents are rather similar to non-respondents and that a significant effect of the trend variable would indicate that respondents with lower levels of employee well-being may be more likely to drop out of the study. We ran the analyses for both a larger sample, including respondents who did not have contact with COVID-19 patients ($n = 1{,}286$) to increase statistical power to detect an effect, and for the main sample of employees who had contact with COVID-19 patients ($n = 468$). We find that the coefficient for the trend variable is very close to zero and non-significant ($p > .17$) regarding all five dependent variables across both samples. We interpret this result as a sign for a lack of a non-respondent bias.

Third, we take into account that the error terms are likely to be correlated across equations (e.g., the error terms of the equations with physical and mental health as dependent variables are potentially correlated). Our current estimation approach ignores this correlation, but we may even exploit the correlation to improve the efficiency of our estimators by applying seemingly unrelated regressions (Cameron, 2010). Seemingly unrelated regressions have two main advantages; First, they take into account potential correlations between the error terms across equations. Second, they allow for testing the joint significance of variables across equations. In the first step, we run seemingly unrelated regressions to analyze our DID estimator for each of the five dependent variables. We found a significant effect of our three-way interaction effect (i.e., crisis x crisis intensity x empowering leadership) on vigor (b = -.004; p < .001), dedication (b = -.003; p < .001), physical health (b = .002; p = .029) and mental health (b = .003; p < .001) and a non-significant effect on absorption (b = .000; p = .806). These findings support our initial results except for the significant effect of dedication. Also, we tested the joint significance of the three-way interaction across the five equations. That is, we analyzed whether the joint effect of the three-way interaction differs from 0 for all five dependent variables. The χ^2 test is significant (χ^2 = 39.34, df = 5; p < .001), which indicates a joint effect of the three-way interaction on our dependent variables.

Fourth, whereas we followed the recommendations by Bertrand and colleagues (2004) and clustered our standard errors at the province level, the standard errors may be biased due to the low number of provinces (k = 12). Therefore, we followed recommendations by Cameron and Miller (2015) and use the cluster bootstrap method to test the significance of the three-way interaction effect. Specifically, we use the *boottest* command in Stata 16.0 with Webb weights and 4,999 replications. We find that the p-values were in most cases slightly higher than in our initial analyses (vigor: p = .037; dedication: p = .224; absorption: p = .855; physical exhaustion: p = .106; mental exhaustion: p = .097). Yet, our main conclusions are supported.

Fifth, a crucial assumption of our DID design is that the COVID-19 hospitalization within a province is a valid proxy for healthcare employees' job demands during the pandemic and, thus, indicates the crisis intensity. Although we provide evidence for the plausibility of this assumption, it is difficult to directly test this assumption. However, we can indirectly test this assumption by analyzing a group of employees who experienced the COVID-19 pandemic, but whose jobs were largely unaffected by the pandemic. In that case, we would expect to identify very little differences between employees in provinces with low and high crisis intensity. We ran all regressions again this time using only employees who have had no contact with COVID-19 patients (n = 818; total number of observations = 1,636). Indeed, we found that despite the larger sample size the three-way interaction effect (crisis x crisis intensity x empowering leadership) was non-significant ($p \geq .26$) regarding all five dependent variables. Additionally, the estimate coefficients were in most cases much smaller (≤ 26%) than the estimated coefficients in our initial models. The results provide evidence that our crisis intensity measure is a valid proxy of the job demands during the COVID-19 pandemic.

A

Appendix to chapter 5

The dataset and syntax can be found here:
https://osf.io/ekdr6/?view_only=f802587d0ff3434cabc1a9a80c42a7eb

Appendix to chapter 6

- **Appendix 1: Dutch translations of items**

Beneath, the items for PSM, PM and intentions to report wrongdoings are presented. All items are measured on a Likert scale ranging from 1 (totally disagree) to 5 (totally agree). The first table presents the values and their meaning.

Table 1 *Likert scale values for all variables*

Value	Meaning	Dutch translation presented
1	Totally disagree	Helemaal mee oneens
2	Disagree	Mee oneens
3	Neither agree nor disagree	Niet mee eens, niet mee oneens
4	Agree	Mee eens
5	Totally agree	Helemaal mee eens

Table 2 *Items for Public Service Motivation. Based on Vandenabeele and Jager (2020)*

Item nr.	Item in English	Dutch translation presented
	Question: Answer the following statements.	Vraag: beantwoord de volgende stellingen.
1	I am very motivated to contribute to society.	Ik ben heel gemotiveerd een bijdrage te leveren aan de maatschappij.
2	Making a difference in society, no matter how small, is very important to me.	Het verschil maken - hoe klein dan ook - in de samenleving vind ik heel belangrijk.
3	Defending the public interest is very important to me.	Ik vind het belangrijk het algemeen belang te verdedigen.
4	I find it to be very motivating being able to contribute to society.	Ik ervaar het als zeer motiverend een bijdrage te kunnen leveren aan de maatschappij.

A

Table 3 *Items for Prosocial motivation. Based on Grant (2008b)*

Item nr.	Item in English	Dutch translation presented
	Question: Why are you motivated to do your job?	Vraag: Waarom ben je gemotiveerd om je werk te doen?
1	I care about benefiting others through my work.	Omdat ik het belangrijk vind dat anderen baat hebben bij mijn werk.
2	I want to help others through my work.	Omdat ik anderen door mijn werk wil helpen.
3	I want to have a positive impact on others through my work.	Omdat ik een positieve impact op anderen wil hebben.
4	It is important for me to do good for others through my work.	Omdat het belangrijk voor me is om het goede te doen door mijn werk.

Table 4 *Items for intentions to report wrongdoings from colleagues and patients. Based on Meyer-Sahling et al. (2019)*

Item in English	Dutch translation presented
I feel comfortable reporting ethical problems to upper management in case of undesirable behaviour from a colleague.	Ik voel me voldoende op mijn gemak om ethische problemen te melden aan mijn leidinggevenden bij ongewenst gedrag van een collega.
I feel comfortable reporting ethical problems to upper management in case of undesirable behaviour from a patient/client.	Ik voel me voldoende op mijn gemak om ethische problemen te melden aan mijn leidinggevenden bij ongewenst gedrag van een patiënt / cliënt.

Appendices to chapter 7

Next to the appendices below, the dataset and syntax can be found here: https://osf.io/6n2g4/?view_only=895be1c46d384867b52e22ff30892ba8

A

- **Appendix 1: Interview guide**

Context: Qualitative pre-study for quasi-field experiment.
Goal pre-study: Developing appropriate nudges to reduce work pressure and stress.
Specific goals: 1) Defining the behavioral problem. 2) Deciding whether choice architecture is appropriate. 3) Checking whether there are bottlenecks that may hamper intervention power.

1. Before the interview
   a Written informed consent to interview participation

2. Introduction

   a Consent to record interview
   b Introducing the researcher
   c Explaining the research
   d Introducing the interviewee
   i Who are you and what is your job title?
   ii How long have you been doing this?

3. Behavioral problem

   a Do you have to deal with work pressure/stress in your work?
   i In your own work?
   ii In work of colleagues/other employees?
   iii What groups of colleagues?
   b In what concrete moments do you experience work pressure/stress? E.g., [partly inductively/deductively generated list]
   i Working overtime
   ii Working extra shifts
   iii Presenteeism
   iv Not taking holidays
   v Taking no breaks
   vi Checking your work phone at home
   vii Not being able to say no to requests (to which requests?)
   viii Using email
   ix Meetings
   c What are the consequences hereof for you?
   i For your work with clients

ii *For your work with colleagues*
iii *For your private life*
d In what concrete situations do you observe work pressure/stress among your colleagues and what are the consequences hereof? For which groups of colleagues?

4. Goal behavior

a How would you or your colleagues behave ideally regarding the situations that you mentioned?

5. Why is there a problem

a What causes the gap between the goal behavior and the behavioral problem?
i *Causes*
ii *Causers*
b Do you have ideas to reduce work pressure/stress regarding this situation?

6. Closing

A

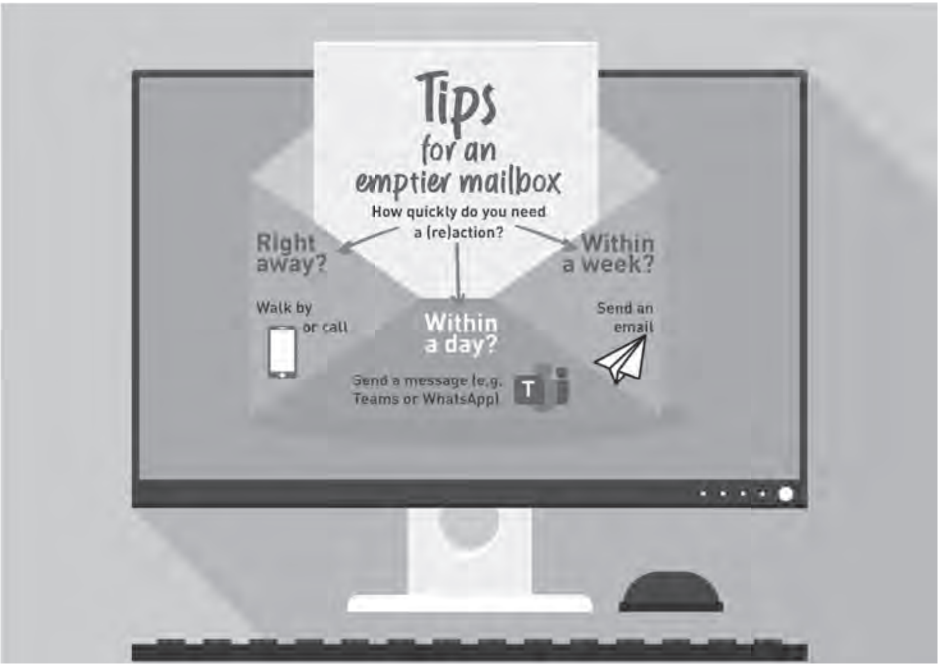
- **Appendix 2: Nudges.**

The first nudge, an opinion leader nudge, is a classic nudge that nevertheless is expected to preserve autonomy because it targets System 2 thinking by providing information. An opinion leader nudge is a message in which the behavior of a person of influence is described. Its expected effect is that receivers of this message will adapt this behavior due to the position of the opinion leader. The mechanism is based on two assumptions. First, in making decisions people rely on social reference points, i.e., the behavior of others (Münscher et al., 2016). For example, descriptive social norms, statements about what other people would do in a situation, stimulate people to conform. Furthermore, opinion leaders are highly respected messengers whose views or behaviors people are more likely to adapt. There are multiple ways of identifying such opinion leaders, but a common way to do so is by selecting those in formal leadership positions or with a specific expertise (Valente and Pumpuang, 2007). The opinion leader nudge in this study presented a message from 'your HR manager' saying that they notice emailing too much causes unnecessary stress and they are therefore going to email less, suggesting this will give more calmness in work, and asking to join them.

Figure 1 *Opinion leader nudge*

The second nudge, a rule-of-thumb, attempts to implement more conscious behavior change, resembling a boost. A rule-of-thumb nudge presents information in an understandable way so that less effort is required to make a decision (Hertwig and Grüne-Yanoff, 2017; Münscher et al., 2016). It is a hybrid behavioral intervention. Münscher et al. (2016) refer to a rule-of-thumb as a simplification nudge, but it also resembles a simple decision tree, which is a type of boost (Hertwig and Grüne-Yanoff, 2017, p. 979). The rule-of-thumb in this study provides a simple question to decide whether email is appropriate: how quickly do you need a (re)action? The suggested rule-of-thumb is that if you need it right away, you should walk by or call; if you need it within a day, you should send a message; and if you need it within a week, you should send an email.

Figure 2 *Rule-of-thumb*



Finally, we describe self-nudges, nudges to be used by employees themselves. The concept of self-nudging was recently introduced by Reijula and Hertwig (2022). Many nudges could be turned into self-nudges, the difference being that the person who is nudged is also the one who nudges. Enabling people to apply self-nudges can be seen as a type of self-control boost (Hertwig and Grüne-Yanoff, 2017, p. 979). Examples

A

of self-nudges reminding yourself of a certain decision (e.g., by putting up a note on their computer screen) or adapting a different frame for the same decision (e.g., by thinking about working out at the gym as a privilege rather than a chore) (Reijula and Hertwig, 2022). In this study, three self-nudges were proposed. To create awareness among employees that they can influence their own choice architecture to change their behavior (Reijula and Hertwig, 2022), we introduced the self-nudges by saying 'You can help making your and your colleagues' mailbox emptier. What challenge do you recognize?' After this, three challenges were introduced, each connected to one of the self-nudges. The challenges are: email response uncertainty ('When I don't know if I should respond to an email, I do it anyway'), real time emailing ('When I have question, I email it directly'), and email addressee uncertainty ('When I don't know who to email, I email everyone'). The three proposed self-nudges are (1) providing your colleagues a timely reminder about whether they need to respond to an email or not (hereby also indirectly reminding yourself of this behavior), (2) providing reminders to yourself about the question you have so you can delay action and have time to think about alternative strategies, (3) reframing emailing to be about the receiver rather than the sender: considering the consequences of limiting the number of addressees for colleagues (reduced stress) rather than for yourself (it may take a bit more time before you receive the right answer).

Figure 3 *Self-nudges, image with behavioral challenges*

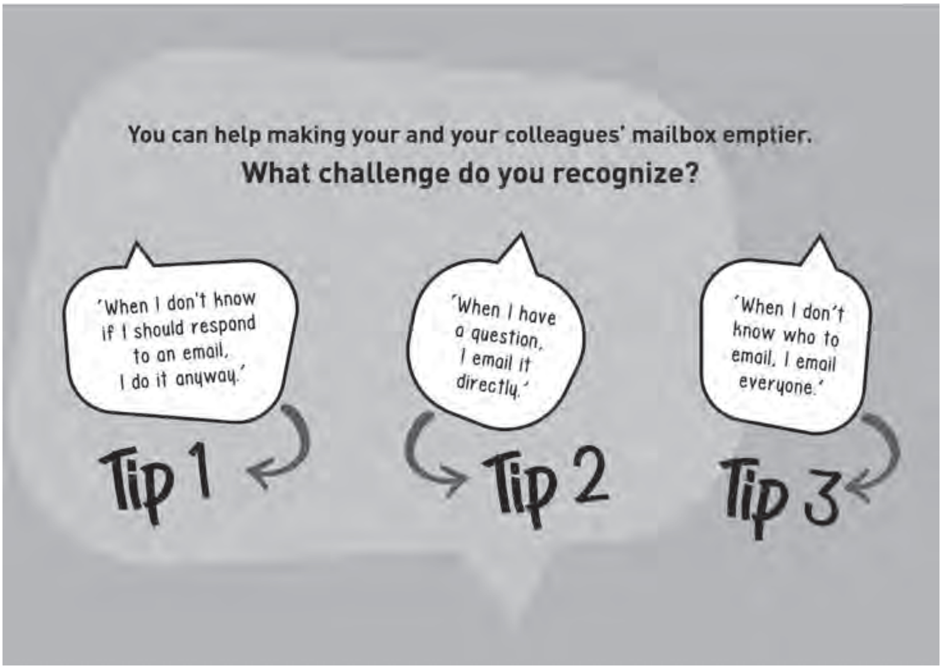
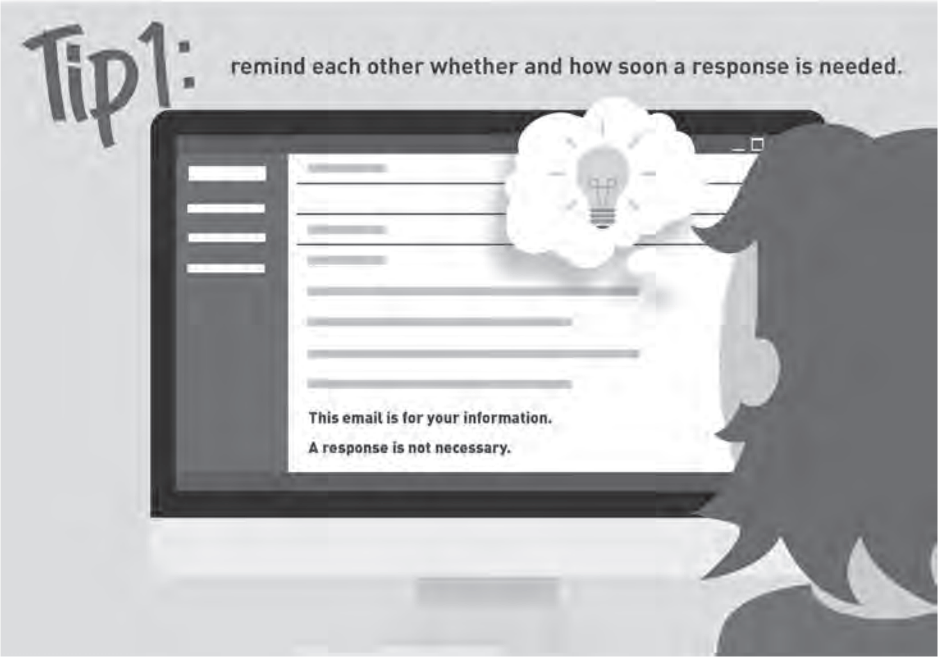


Figure 4 *Self-nudges, self-nudge 1*



A

Figure 5 *Self-nudges, self-nudge 2*

Figure 6 *Self-nudges, self-nudge 3*

A

- **Appendix 3: Survey measures**

This appendix introduces the survey measures, including an elaborate explanation of the Bayesian Truth Serum. In the survey experiment, questions were translated to Dutch.

Email volume (Sumecki et al., 2011)
Measured with two open questions with numerical content validation:
You've been on holiday for one week. How many new emails would you expect to find in your inbox when you return to work? (in digits)

On an average working day, how much time (in minutes) do you spend managing emails (reading, sending, filing, etc.)? (in digits)

Email overload (Dabbish and Kraut, 2006)
Measured with seven items on a 7-point Likert scale ranging from 'strongly disagree' to 'strongly agree':
1: *I can handle my email efficiently.* (Reversed)
2: *I have trouble finding information in my email.*
3: *I can easily deal with the amount of email I receive.* (Reversed)
4: *I sometimes miss information or important email messages.*
5: *I reply quickly to the email message I need to.* (Reversed)
6: *Dealing with my email disrupts my ongoing work.*
7: *I find dealing with my email overwhelming.*

Feasibility, appropriateness, meaningfulness and effectiveness (FAME-approach for evidence-based practice, Jordan et al., 2019)
Measured with multiple separate 7-point Likert scales:
How feasible would it be to use this message in your organization? ('Very unfeasible' to 'Very feasible')

How appropriate would this message be in your organization? ('Very inappropriate' to 'Very appropriate')

How meaningful would this message be to your organization? ('Very meaningless' to 'Very meaningful')

How effective would this message be in decreasing email use in your organization? ('Very ineffective' to 'Very effective')

Bayesian Truth Serum for non-compliance (John et al., 2012; Prelec, 2004)
The Bayesian truth serum is a scoring algorithm that combines the answers of respondents about their behavior and their estimates of what others would answer (John et al., 2012, p. 526). The serum increases credibility by developing multiple estimates (elaborated below) rather than one estimate about respondents' own behavior (i.e., would you comply) and combining these estimates into a conservative judgement of, in our case, non-compliance. The serum has been used in large-scale surveys (Van de Schoot et al., 2021; Frank et al., 2017; Weaver and Prelec, 2013) and scholars recommend using this approach in experimental social science research (Schoenegger, 2023).

In our setup, it worked as follows. First, we briefly explained its purpose to respondents: 'we would like to ask you to predict the response of your colleagues and yourself. The following questions help us predict the effect of this message'. We then asked respondents to estimate three values: 1) the percentage of colleagues that would send as many emails after the intervention as before (the prevalence estimate), 2) the percentage of colleagues that would be honest about sending as many emails after the intervention as before (the admission estimate), and 3) whether they would send less emails after the intervention (yes or no; the self-admission rate). This method allows us to report three different estimates of non-compliance: self-admission rates, prevalence estimates, and prevalence estimates calculated by dividing the self-admission rates by the admission estimates. We generate a more conservative judgement of non-compliance by taking the geometric mean of these three values (John et al., 2012) as indicated in formula 1. Subtracting the proportion of non-compliance from 100 gives us the proportion of compliance. In our results, we will compare the self-admission estimates to the geometric mean.

$$(1) \qquad \sqrt[3]{Self\ admission\ rate\ in\ \% * Prevalence\ estimate * \left(\frac{Self\ admission\ rate\ in\ \%}{Admission\ estimate}\right)}$$

The specific questions are presented below:
Introduction: *We would like to ask you to predict the response of your colleagues and yourself. The following questions help us predict the effect of this message.*

1: *What percentage of your colleagues would send as many emails after this message as before?* (from 0% to 100% using a graphic slider) (prevalence estimate)

2: *What percentage of your colleagues would be honest about sending as many emails after this message as before?* (from 0% to 100% using a graphic slider) (admission estimate)

3: *Are you going to send less emails after this message?* (Yes/No) (self-admission rate)

Perceived autonomy (Morgeson and Humphrey, 2006; Gorgievski et al., 2016)
Measured with three items on a 7-point Likert scale ranging from 'strongly disagree' to 'strongly agree':
1: *This message gives me a chance to use my personal judgement in using email.*
2: *This message allows me to make a lot of decisions about using email on my own.*
3: *This message provides me with significant autonomy in making decisions about using email.*

Work engagement (UWES-3, Schaufeli et al., 2019)
Measured with a 5-point Likert scale ranging from 'Never' (1) to 'Always (daily)' (5):
1: *At my work, I feel bursting with energy.*
2: *I am enthusiastic about my job.*
3: *I am immersed in my work.*

- **Appendix 4: Methods and results pilot study**

This appendix describes the methods and results of the pilot study.

Methods

Participants

We collected data on March 22 from 435 respondents via Prolific in an English language Qualtrics survey. The sample size was determined with an *a priori* power analysis using G*Power. The survey did not inquire about personal data except those provided by Prolific itself, for which respondents provided consent. Respondents were required to work full-time and use email in their job. The mean age of the 435 respondents was 32.75 (SD = 8.71, Min. = 19, Max. = 69). Regarding gender, 204 were female, 230 were male and 1 respondent indicated they would rather not say.

Procedure and measures

Respondents first provided background characteristics, including an eligibility check (respondents had to work full-time and use email at their job). We assessed email volume and email time with open questions adapted from Sumecki et al., 2011 (p. 409): 'You've been on holiday for one week. How many new emails would you expect to find in your inbox when you return to work? (in digits)' and 'On an average working day, how much time (in minutes) do you spend managing emails (reading, sending, filing, etc.)? (in digits)'. We also assessed email overload with 7 items on a 7-point Likert scale (α = .82) ranging from 'strongly disagree' to 'strongly agree' (Dabbish and Kraut, 2006). Next, respondents were exposed to one of the three nudges randomly. Chi-square tests indicated that randomization was successful among gender (male versus female) and age groups (younger versus older than mean age) as no significant differences existed (χ^2 (2) = .42, p = .813 for gender and χ^2 (2) = .13, p = .936 for age). The instruction accompanying the nudges read 'Imagine the organization you work for sends you the following message about using email in your organization. (If you fill out this survey on a mobile phone, you are able to zoom in.) Please read the message carefully.' After the nudge, they were asked to assess perceived autonomy on a 7-point Likert scale (α = .86), a scale we created by adapting the three item Decision-Making Autonomy subscale from the Work Design Questionnaire (WDQ; Morgeson and Humphrey, 2006). Inspired by Jordan et al. (2019), we also assessed the feasibility, appropriateness, meaningfulness and expected effectiveness of the nudge with single items on 7-point Likert scales, asking respondents to imagine the effects of this message in their organization.

To come to our final sample (n = 435), listwise deletion was applied and respondents had to pass an attention check (a multiple-choice question with two boxes and the instruction 'please only check box 2'). Three respondents failed this check. We use one way analysis of variance for our main analysis and Kruskal-Wallis H Tests for additional analyses (these were used as we used ordinal variables. For interpretation, means and standard deviations were included). Significance levels were set at $p = .05$ (for all models, exact p-levels were reported).

Results

Table 1 present the means, standard deviations and correlations of the main variables.

Table 1 *Correlations. n = 435*

	M (SD)	1	2	3	4	5	6	7	8
1. Email volume	106.88 (257.23)	-	-	-	-	-	-	-	-
2. Email time	73.24 (78.52)	.35*	-	-	-	-	-	-	-
3. Email overload	2.64 (1.03)	.26*	.20*	-	-	-	-	-	-
4. Autonomy	5.26 (1.18)	.03	.03	-.06	-	-	-	-	-
5. Feasibility	4.56 (1.65)	.003	-.001	.02	.31*	-	-	-	-
6. Appropriateness	4.86 (1.50)	.07	.07	.06	.41*	.67*	-	-	-
7. Meaningfulness	4.58 (1.54)	.08	.05	.07	.47*	.67*	.69*	-	-
8. Effectiveness	4.11 (1.58)	-.03	.00	.008	.43*	.56*	.54*	.71*	-

Note. * $p < .001$ (two-tailed). Correlations are Pearson except for those with email volume and email time, these are Spearman as for these variables the data indicated outliers.

A one-way analysis of variance showed that the effects of the nudges on perceived autonomy differed significantly, $F(2,432) = 5.21$, $p = .006$ ($\eta^2 = .024$)[1]. Tukey HSD post hoc analyses indicated that the perceived autonomy was significantly lower for the rule-of-thumb ($M = 5.01$, $SD = 1.26$) compared to the opinion leader nudge ($M = 5.41$, $SD = 1.15$) ($p = .009$) and the self-nudges ($M = 5.37$, $SD = 1.07$) ($p = .026$). Perceived autonomy did not differ significantly across age, gender or email overload (t-test with dummy variable younger/older than the mean age: $t(433) = .42$, $p = .672$[2]; t-test with dummy variable male/female: $t(432) = -1.77$, $p = .078$[3]; t-test with dummy variable below mean email overload/above mean email overload: $t(431.46) = -.39$, $p = .699$[4]).

1 Equal variances assumed as $F(2,432) = 2.06$, $p = .129$.
2 Two-sided p. Equal variances assumed as $F(2,432) = 2.81$, $p = .094$.
3 Two-sided p. Equal variances assumed as $F(2,432) = .31$, $p = .577$.
4 Two-sided p. Equal variances not assumed as $F(2,432) = 5.08$, $p = .025$.

We conducted Kruskal-Wallis H Tests and found that effects of the nudges differed on perceived feasibility, appropriateness and meaningfulness, but not effectiveness (Table 2). Pairwise comparisons indicate that for feasibility, the opinion leader nudge scored significantly ($p < .05$) lower than the other nudges. For appropriateness, the self-nudges scored significantly higher than the other nudges. For meaningfulness, the self-nudges scored significantly higher than the opinion leader nudge.

Table 2 *Kruskal-Wallis H Tests and descriptives*

	Feasibility	Appropriateness	Meaningfulness	Effectiveness
Test result	$H(2) = 19.49$, $p < .001$	$H(2) = 17.62$, $p < .001$	$H(2) = 6.32$, $p = .042$	$H(2) = 4.35$, $p = .114$
Means (*SD*s) and pairwise comparisons				
Opinion leader	4.11 (1.69)^{ab}	4.55 (1.55)^{a}	4.38 (1.62)^{a}	4.03 (1.63)
Rule-of-thumb	4.63 (1.63)^{a}	4.75 (1.51)^{b}	4.53 (1.52)	3.97 (1.65)
Self-nudges	4.94 (1.52)^{b}	5.28 (1.35)^{ab}	4.83 (1.47)^{a}	4.35 (1.45)

Note. ^{a}and ^{b}: Categories that have significantly differing mean rank scores cf. the Kruskal-Wallis H Tests ($p < .05$). Significance values were adjusted with the Bonferroni correction for multiple tests.

- **Appendix 5: Traditional email interventions.**

The traditional interventions were based on real strategies that organizations have to reduce email stress, resembling traditional policy instruments (Tummers, 2019). The first intervention is technological. Organizations could limit email access to specific hours in a day so that employees are limited to checking email a few times a day. We term this the 'email access limit'. It resembles the 'whip' approach (Tummers, 2019) as well as organizational strategies to force employees to take time off. The next intervention is an economic incentive and resembles the 'carrot' approach (Tummers, 2019). Organizations could reward employees financially for reducing their email use. We term this the 'monetary reward'. Offering employees monetary rewards for meeting goals is a very common business practice (Aguinis et al., 2013). The last intervention is variation on the 'carrot' by means of a social incentive. In the intervention 'public praise', employees are publicly praised for showing exemplary behavior in email use. Research suggests public praise may be more effective than financial incentives (Handgraaf et al., 2013).

Table 1 *Traditional interventions*

Intervention	Dutch text	English translation
Email access limit	Binnen onze organisatie wordt veel gemaild. Dit kan zorgen voor stress. Daarom kunnen medewerkers vanaf nu alleen tussen 10 en 11 uur ’s ochtends en 3 en 4 uur ‘s middags e-mail versturen of ontvangen. We hopen dat dit helpt om je een legere mailbox te bezorgen	There is a lot of emailing within our organization. This can cause stress. Therefore, from now on, employees can only send or receive e-mail between 10 a.m. and 11 a.m. and 3 p.m. and 4 p.m. We hope this helps to give you an empty mailbox.
Monetary reward	Binnen onze organisatie wordt veel gemaild. Dit kan zorgen voor stress. Daarom hebben we uitgerekend hoeveel e-mails jij gemiddeld per werkdag verstuurt. Vanaf nu krijg je per werkdag voor elke e-mail die je minder stuurt dan dit gemiddelde, 1 euro extra bij je volgende salarisstrook. We hopen dat dit helpt om je een legere mailbox te bezorgen.	There is a lot of emailing within our organization. This can cause stress. That is why we have calculated how many e-mails you send on average per working day. From now on, per working day you will receive 1 euro extra for every e-mail that you send less than this average. We hope this helps to give you an empty mailbox.
Public praise	Binnen onze organisatie wordt veel gemaild. Dit kan zorgen voor stress. Daarom sturen we vanaf nu elke week een lijst met ‘minder e-mail helden’ rond: dit zijn de medewerkers die binnen hun functie de minste e-mails hebben verstuurd. We hopen dat dit helpt om je een legere mailbox te bezorgen.	There is a lot of emailing within our organization. This can cause stress. That is why from now on we will send out a list of 'less-email heroes' every week: these are the employees who have sent the fewest emails within their position. We hope this helps to give you an empty mailbox.

A

- **Appendix 6: Randomization survey experiment**

To assess whether randomization between the different interventions in the survey experiment was successful, we computed multiple chi-square tests for independence.

Gender

Table 1 presents the division of gender across interventions. A Chi-square test indicated that randomization was successful among gender (male or female): $\chi^2(6) = 2.48$, $p = 0.871$.

Table 1 *Gender across interventions*

Intervention	Gender		Total
	Female	Male	
Opinion leader	510	73	583
Rule-of-thumb	509	83	592
Self-nudges	491	88	578
All nudges	493	84	577
Email access limit	506	83	589
Monetary reward	496	89	585
Public praise	501	87	588
Total	3506	587	4093

Note. Respondents reporting X or that they would rather not say ($n = 19$) were left out of this analysis.

Age groups

Table 2 presents the division of age groups across interventions. A Chi-square test indicated that randomization was successful among age groups: $\chi^2(30) = 32.29$, $p = 0.354$.

Table 2 *Age groups across interventions*

Intervention	Age groups						Total
	25 or younger	26-35	36-45	46-55	56-65	66 or older	
Opinion leader	6	46	102	169	258	7	588
Rule-of-thumb	3	46	90	180	265	7	591
Self-nudges	6	35	97	186	252	8	584
All nudges	7	30	89	182	261	9	578
Email access limit	3	34	116	195	238	4	590
Monetary reward	3	34	131	164	248	7	587
Public praise	3	34	100	190	255	9	591
Total	31	259	725	1266	1777	51	4109

Note. A total of 3 respondents did not disclose age.

Healthcare sector

Table 3 presents the division of healthcare sectors across interventions. A Chi-square test indicated that randomization was successful among healthcare sector: $\chi^2(24) = 24.25$, $p = 0.447$.

Table 3 *Healthcare sectors across interventions*

Intervention	Healthcare sectors					Total
	Hospitals	Nursing/home care	Mental healthcare	Disabled care	Other	
Opinion leader	221	156	101	79	32	589
Rule-of-thumb	226	156	74	91	45	592
Self-nudges	213	142	99	97	33	584
All nudges	231	140	92	85	30	578
Email access limit	199	152	98	90	52	591
Monetary reward	206	155	91	95	40	587
Public praise	219	158	98	83	33	591
Total	1515	1059	653	620	265	4112

Working hours

Table 4 presents the division of working hours across interventions A Chi-square test indicated that randomization was successful among amount of working hours per week (29 hours or more versus less than 29): $\chi^2(6) = 2.589$, $p = 0.858$.

Table 4 *Working hours across interventions*

Intervention	Working hours		
	Less than 29	**29 or more**	**Total**
Opinion leader	282	304	586
Rule-of-thumb	281	309	590
Self-nudges	281	303	584
All nudges	270	306	576
Email access limit	300	290	590
Monetary reward	278	307	585
Public praise	291	297	588
Total	1983	2116	4099

Note. Respondents who reported to have a zero-hours contract (n = 13) were left out of this analysis.

- **Appendix 7: Preregistration and evaluation of original hypotheses**

Preregistration and deviations

Initially, the pilot study, the survey experiment and the quasi-field experiment were all preregistered separately with a total of four hypotheses (See List of preregistrations). In this paper, we combined the initial preregistrations into one main hypothesis. Below we explain how, and we briefly discuss all original hypotheses.

Specifically, we moved the pilot study to an appendix. The hypothesis of the pilot study focused on the slight differences in perceived autonomy between the nudges, but we decided that the main paper should rather focus on the autonomy and effectiveness of nudges in general. Second, to come to our one main hypothesis, we merged all three remaining hypotheses from the survey experiment and quasi-field experiment. The original hypotheses separately addressed nudges 1) preserving autonomy more than traditional interventions, 2) being perceived as less effective than traditional interventions, 3) being effective in decreasing email use. We decided to introduce the traditional interventions in the method section and not explicitly mention them in the theory section. Therefore, our main hypothesis in the paper focuses on the expectation that nudges are autonomy-preserving and effective in decreasing email use. In our method section, we now describe how we test this by evaluating the absolute and relative scores (the latter in comparison to traditional interventions) of autonomy and nudge effectiveness, and by testing both subjective and objective nudge effectiveness.

Regarding the preregistration of analyses, one major deviation should be mentioned. The quasi-field experiment was analyzed with a slightly different statistical analysis, linear mixed models rather than ANOVA. The preregistered analyses appeared less suitable for analysis when the data came in, and linear mixed models allowed for more flexibility. Specifically, this analysis allowed for fixed factors to be included, in our analysis time (the week) was the repeated measure fixed factor (Krueger and Tian, 2004). The change of analysis did not fundamentally change our results.

Evaluation of original hypotheses

Pilot study

We expected self-nudges to be most autonomy-preserving, as it combined full transparency with the ability to influence one's own choice environment. We expected an opinion leader nudge to be the least autonomy-preserving, as it employs hierarchy (i.e., the opinion leader) to make employees change behavior. The rule-of-thumb was estimated to score in between the self-nudges and the opinion leader nudge, as it does not employ hierarchy but does suggest a specific behavior change.

H1: *A rule-of-thumb nudge will preserve autonomy more than an opinion leader nudge (a) but less than self-nudges (b).*

The results in Appendix 3 show that perceived autonomy was significantly lower for the rule-of-thumb in comparison to the other nudges. Therefore, the first part (a) of the above hypothesis was not confirmed, but the second part (b) was. A possible explanation to why the rule-of-thumb was considered the least autonomy-preserving, may be that this nudge was more specific in telling employees how to behave. What is more, employees may not have experienced the opinion leader nudge as imposing hierarchical pressure. The implicit hypothesis underlying this hypothesis, namely that nudges will preserve autonomy, is evaluated in the main paper.

Survey study

Two hypotheses were preregistered for the survey study. They both specifically compare nudges to the traditional interventions. In the final paper, we decided to move the introduction of the traditional interventions to the method section, and have it be only part of the evaluation of the hypothesis besides evaluating the absolute scores on perceived autonomy and effectiveness. Nevertheless, the first hypothesis below was confirmed by our results: nudges preserve autonomy more than traditional interventions. The second hypothesis, however, was not: we expected employees may predict that the traditional interventions would be more effective as they offered more rigorous (yet less autonomy-preserving) ways to change behavior. In contrast to this expectation, employees thought the nudges would be more effective.

H1: *Nudges will preserve autonomy more than technical, social or economic interventions.*

H2: *Employees will predict that technical, social or economic interventions are more effective than nudges.*

Quasi-experiment

The original hypothesis for the quasi-experiment stated nudges would be effective in decreasing actual email use and is still evaluated as part of the combined hypothesis in the paper. While we observe decreases in email use, the partially insignificant results suggest we should be careful in our conclusions.

H1: *An opinion leadership nudge, rule-of-thumb and self-nudges will decrease email volume.*

Co-author statements

Co-author statement for Utrecht School of Governance dissertations

Universiteit Utrecht

Contribution

This co-author statement regards the following contribution:
Van Roekel, H., van der Fels, I. M., Bakker, A. B., & Tummers, L. G. (2021). Healthcare workers who work with COVID-19 patients are more physically exhausted and have more sleep problems. *Frontiers in Psychology, 11.* DOI: 10.3389/fpsyg.2020.625626.

Author roles and contributions

Contributor Roles Taxonomy (CRediT) (source: https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement; see the source for an explanation of concepts)

	Henrico van Roekel	Irene van der Fels	Arnold Bakker	Lars Tummers
Project coordination	X			
Conceptualization	X	X	X	X
Research design	X	X	X	X
Privacy and ethics approval	X	X		
Data collection	X	X		
Data analysis	X	X		
Data curation	X			
Writing - original draft	X			
Writing - review & editing	X	X	X	X

Signatures authors, including signature PhD candidate

Date	Name	Signature
30-11-23	Henrico van Roekel	
27-11-23	Irene van der Fels	
29-11-23	Arnold Bakker	
28-11-23	Lars Tummers	

Co-author statement for Utrecht School of Governance dissertations

Contribution

This co-author statement regards the following contribution:

Van Roekel, H., Wigger, E. F. J., Veldkamp, B. P., Bakker, A. B. (2023). What is work engagement? A text mining approach using employees' self-narratives. *Applied Psychology: An International Review.* DOI: 10.1111/apps.12501.

Author roles and contributions

Contributor Roles Taxonomy (CRediT) (source: https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement; see the source for an explanation of concepts)

	Henrico van Roekel	**Enno Wigger**	**Bernard Veldkamp**	**Arnold Bakker**
Project coordination	X			
Conceptualization	X	X	X	X
Research design	X	X	X	X
Privacy and ethics approval	X			
Data collection	X			
Data analysis	X	X		
Data curation	X	X		
Writing - original draft	X	X		
Writing - review & editing	X	X	X	X

Additional remarks:

Henrico and Enno shared first authorship.

Signatures authors, including signature PhD candidate

Date	**Name**	**Signature**
30-11-23	Henrico van Roekel	
27-11-23	Enno Wigger	
27-11-23	Bernard Veldkamp	
29-11-23	Arnold Bakker	

C

Co-author statement for Utrecht School of Governance dissertations

Universiteit Utrecht

Contribution

This co-author statement regards the following contribution:
Van Roekel, H., Sieweke, J., Schott, C., Bakker, A. B., & Tummers, L. G. The effect of empowering leadership on employee well-being during a public health crisis: a natural experiment. *Under review.*

Author roles and contributions

Contributor Roles Taxonomy (CRediT) (source: https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement; see the source for an explanation of concepts)

	Henrico van Roekel	Jost Sieweke	Carina Schott	Arnold Bakker	Lars Tummers
Project coordination	X				
Conceptualization	X	X	X	X	X
Research design	X	X	X	X	X
Privacy and ethics approval	X				
Data collection	X				
Data analysis	X	X			
Data curation	X				
Writing - original draft	X	X			
Writing - review & editing	X	X	X	X	X

Signatures authors, including signature PhD candidate

Date	Name	Signature
30-11-23	Henrico van Roekel	
27-11-23	Jost Sieweke	
27-11-23	Carina Schott	
29-11-23	Arnold Bakker	
28-11-23	Lars Tummers	

Co-author statement for Utrecht School of Governance dissertations

Contribution

This co-author statement regards the following contribution:
Van Roekel, H., & Schott, C. (2022). Activating employees' motivation to increase intentions to report wrongdoings: evidence from a large-scale survey experiment . *Public Management Review*. DOI: 10.1080/14719037.2021.2015184.

Author roles and contributions

Contributor Roles Taxonomy (CRediT) (source: https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement; see the source for an explanation of concepts)

	Henrico van Roekel	**Carina Schott**
Project coordination	X	
Conceptualization	X	X
Research design	X	X
Privacy and ethics approval	X	
Data collection	X	
Data analysis	X	
Data curation	X	
Writing - original draft	X	X
Writing - review & editing	X	X

Signatures authors, including signature PhD candidate

Date	**Name**	**Signature**
30-11-23	Henrico van Roekel	
27-11-23	Carina Schott	

Co-author statement for Utrecht School of Governance dissertations

Universiteit Utrecht

Contribution

This co-author statement regards the following contribution:
Van Roekel, H., Giurge, L., Schott, C., & Tummers, L. (2023). Nudges can be both autonomy-preserving and effective: Evidence from a survey and quasi-field experiment. *Behavioural Public Policy*. DOI: 10.1017/bpp.2023.18.

Author roles and contributions

Contributor Roles Taxonomy (CRediT) (source: https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement; see the source for an explanation of concepts)

	Henrico van Roekel	**Laura Giurge**	**Carina Schott**	**Lars Tummers**
Project coordination	X			
Conceptualization	X	X	X	X
Research design	X	X	X	X
Privacy and ethics approval	X			
Data collection	X			
Data analysis	X			
Data curation	X			
Writing - original draft	X			
Writing - review & editing	X	X	X	X

Signatures authors, including signature PhD candidate

Date	**Name**	**Signature**
30-11-23	Henrico van Roekel	
28-11-23	Laura Giurge	
27-11-23	Carina Schott	
28-11-23	Lars Tummers	

Acknowledgements

Writing this dissertation would not have been possible without the support of important people around me. In this section, I want to thank them specifically.

First, I would like to thank my supervisors: Carina Schott, Lars Tummers, and Arnold Bakker.

As my daily supervisor, Carina, you were always there. In our meetings, we could talk about anything related (or unrelated) to academic life. If I returned from team meetings somewhat overstimulated, you would take the time to chat. But I also learned a lot from you content-wise, like your proficiency in theory building and experience as a teacher. It is safe to say that your role in this process has been indispensable.

Lars, I think your leadership style worked well for me. You were ambitious but always realistic. More importantly, you were approachable and there when it mattered. Not only would you congratulate a success, but you would also call after one of many rejections. Your drive, sincerity, and vision of public administration research have been important to me and will hopefully continue to stimulate me in the years ahead.

Arnold, I have learned a lot from you. I feel you have seen and done everything in the academic world. Your feedback and ideas were always spot-on. Yet, in between serious talks, you did not shy away from sharing anecdotes of everything you encountered. You brought experience, a psychological perspective and a world of research that inspired me.

Many fellow academics and professionals supported this research. I want to thank Irene van der Fels, Enno Wigger, Bernard Veldkamp, Jost Sieweke and Laura Giurge for co-authoring one of the publications in this dissertation. I have learned so much from each of you. Likewise, this research benefitted greatly from all the feedback I received at countless conferences and research days and the valuable lessons that editors and reviewers taught me throughout the peer review process. Also, a thank you to the doctoral committee for the time and effort spent reading this dissertation and participating in the defense. I also want to thank the many people who have helped realize one of the research projects. Finally, thank you to all healthcare employees who provided invaluable insights into their work lives.

I want to thank my colleagues at USBO: Machiel van der Heijden (also for temporarily taking over as daily supervisor!), Pauline Mochel (as BKO tutor!), Mirko Noordegraaf, Tom Overmans, Judith van Erp, Stephan Grimmelikhuijsen, Rosanna Nagtegaal, Paul ‘t Hart, Esther Verheijen, Albert Meijer, Martijn van der Meulen, and many others! Likewise, the support of my fellow PhDs has made the past four years a lot more fun! So, thank you, Lukas, Michelle, Renée, Laura, Anushka, Iris, Luca, Krista, Esther, Kees, Evelien, Isabelle, Lars, Lennart, and all the other (ex-)PhDs. (Gabriela, Isa, and Sheeling, I know you’re looking for your names, but you’ll have to wait…)

I want to thank another group of ‘colleagues’ at Stichting IZZ. You have been very significant to me throughout my PhD trajectory. We have worked together on many relevant topics related to healthcare employee well-being, and it has been a pleasure. You have allowed me to collect lots of data, helped me interpret results throughout many meetings, and generously offered countless ways to disseminate my research. So, thank you, Anouk ten Arve, Irene van der Fels, Marc Spoek, Roland Kip, Cynthia van Wachem, Sonja Lubbers, Lynn Colder, and many others.

I would also like to thank Elizabeth Linos, Jessica Lasky-Fink, and others at Harvard University for having me as a research fellow. My stay in Boston has been an unforgettable experience during my time as a PhD student.

Gabriela, Isa, and Sheeling, as soon as we met, we formed a PhD gang. We have had many adventures together, and I hope many will follow. While some of you may be moving further away, I hope to see much more of you in the future!

Paps en mams, bedankt voor jullie steun en liefde door de jaren. Jullie hebben me altijd gestimuleerd om door te leren en te doen waar ik blij van word. Ik bedank ook graag mijn verdere (schoon)familie. In het bijzonder wil ik mijn zwager Bart en schoonzus Andrea bedanken voor het ontwerpen van dit boek!

Ik wil ook een paragraaf wijden aan mijn vrienden in Utrecht en daarbuiten, voor de steun en welkome afleiding. Alle gesprekken, activiteiten, vakanties, borrels en diners hebben mij zeker geholpen om dit proefschrift af te ronden. Bedankt! Een speciale vermelding hierbij voor mijn paranimfen: Willeke en Evert Jan. Als respectievelijk toegewijde sportbuddy en jarenlange huisgenoot spelen jullie een bijzondere rol.

Joram, de laatste woorden zijn voor jou. Wat ben ik blij dat jij bestaat. Bedankt dat je mijn leven zoveel mooier maakt. Zonder jou had ik dit niet kunnen doen. Ik hoop nog vele jaren met jou door te brengen!

About the author

Henrico van Roekel was born in 1995 in Ede, the Netherlands. In 2016, he obtained a bachelor's degree in public administration and organizational science at the Utrecht University School of Governance (USG). He wrote a bachelor's thesis on the activation of undocumented asylum seekers. He also interned at CNV, one of the largest Dutch trade unions. Next, he completed a two-year research master's program in public administration and organizational science, a joint program of Utrecht University and other Dutch universities. He graduated in 2019 with a master's thesis that presented a field experiment in which the effects of nudging and boosting were tested on the hand hygiene compliance of hospital nurses.

Between 2019 and 2023, Henrico worked as a PhD Candidate at USG in cooperation with Erasmus University Rotterdam and Stichting IZZ. In his PhD project, he focused on improving healthcare employee well-being through leadership, including empowerment and behavioral insights. His research has been presented at multiple (inter)national research conferences and published in journals such as *Public Management Review, Behavioural Public Policy* and *Applied Psychology: An International Review*. During his PhD project, Henrico collaborated with researchers from Oxford University, the London School of Economics, the University of Twente and the Vrije Universiteit Amsterdam. He also taught and coordinated courses at USG and supervised theses of bachelor's and executive master's students.

Furthermore, his research has been featured in (inter)national media outlets like Behavioral Scientist, NRC and BNR, and Henrico has given masterclasses at, among others, the Dutch Ministry of Health, Welfare and Sport, and SPDI, a partnership of Dutch trade unions. As a board member of the *Vereniging voor Bestuurskunde* (the Dutch Association for Public Administration), Henrico aims to make public administration research accessible to the public. In 2022, he was a visiting scholar (and he has been an affiliated scholar since) at The People Lab, Harvard University.

As an assistant professor at USG from September 2023, Henrico aims to continue to study the public sector, its employees and clients through a behavioral lens.

A

WORKING ON WELL-BEING

While healthcare employees provide invaluable contributions to the well-being of patients, their well-being has increasingly been put under pressure. How can we take care of healthcare employees? *Working on Well-being* aims to deepen our understanding of employee well-being in healthcare and develop strategies for leaders to improve the well-being of their employees.

This dissertation first aims to contribute to the literature on well-being by dissecting employees' experiences and innovating employee well-being measurement. Next, it explores leadership as a pivotal force in shaping well-being by studying two contemporary approaches to leadership: empowerment and behavioral insights. It assesses the boundary conditions of empowerment as an effective leadership approach by paying attention to context and employee willingness. Finally, it studies how behavioral insights can be employed rigorously by investigating the mechanisms of effectiveness and elaborate testing.

In studying well-being and leadership, this research presents various research methodologies aimed at innovating survey research. These include text mining, a Bayesian truth serum and self-generated identification codes for longitudinal analysis. It also highlights the use of Open Science practices like ethical review, preregistration and open data.

Through a wide variety of academic insights and practical strategies on well-being and leadership, this dissertation aims to contribute to creating a healthier healthcare system—one in which the well-being of those who dedicate their lives to the well-being of others is also prioritized and protected.