Firms as Architects of Employee Development

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2026

vorgelegt
von

Leonhard Grabe, M.Sc.

aus
Bochum

Referent:	Prof. Dr. Dirk Sliwka
Korreferentin:	Prof. Dr. Judith Künneke
Tag der Promotion:	26.03.2026

Contents

List of Figures

List of Tables

Although this thesis carries the name of only one person, it is the product of years of collaborative work by many researchers and practitioners who have accompanied me throughout this journey. All of them have gone above and beyond their usual duties to ensure the success of the projects related to this thesis. I would like to thank my supervisor Dirk Sliwka for his guidance throughout the process of my doctoral studies. His expertise and active support in our joint projects were very helpful to me. With his genuine interest in practice-oriented research and his appreciation of all employees of this chair, he has always been a role model for me. I would also like to thank my co-supervisor Judith Künneke for her helpful feedback and her commitment to supervising this thesis. I would like to thank Matthias Heinz for his support in our project with the Ministry of Justice and for giving me the opportunity to continue working at the University of Cologne. His reliability and blind trust in my work gave me a sense of security and peace of mind when I needed it the most. My thanks also go to Heiner Schumacher for his support and hospitality during my time at the University of Innsbruck, and to In Gyun Baek for the unique opportunity to broaden my professional and personal horizons at NUS. I am thankful to my colleagues who have accompanied me on my journey at the WiSo Faculty and before – in particular Jakob Alfitian, Sidney Block, Simon Lübke, Saskia Opitz, Johannes Rottmann, Niclas Schmitz, Paula Scholz, Marina Talantcheva, Max Thon, and Patrick Vestner. I would also like to thank Johannes Schultz, Holger Gerhardt, and my other former colleagues at the Center for Economics and Neuroscience in Bonn, who sparked my interest in academic research and encouraged me to pursue a doctoral degree. Finally, I would like to express my deepest gratitude to my parents, my brother Clemens, my partner Jacqueline, and my best friends, Najim and Thilo. Their unwavering belief in my abilities and tireless support carried me through the ups and downs of this thesis.

Introduction

Human capital is a central production factor for firms and has become even more important in periods of rapid technological change. To benefit from technological innovation, firms require employees with appropriate and topical skills. Firms can obtain such skills either by sourcing workers externally or by investing time and resources in training and developing their existing workforce. As jobs become more complex and labor markets exhibit persistent shortages of high-skilled workers, internal development becomes an increasingly attractive option, particularly in settings where firm-specific knowledge, coordination, and learning over time are important (Autor et al., 2003; Bachmann et al., 2022; Lipowski, 2024). A large literature has examined the formation of human capital *outside* the firm, documenting substantial returns to investments in formal education and training, including primary and secondary education, college major choice, and vocational training (Alfonsi et al., 2024; Card, 2001; Field et al., 2019). This work shows that investments in both cognitive and non-cognitive skills can yield substantial individual returns (Deming, 2022; Hanushek et al., 2015; Mincer, 1974). However, these individual investment incentives are not always aligned with the interests of a particular firm (Acemoglu & Pischke, 1998; Becker, 1964), and firms need to actively align the employees' development efforts with their overall strategy. To take a more active role in shaping employee development, firms can draw on a range of organizational control mechanisms. These instruments are commonly referred to as *management practices* or *management controls* and include monetary and non-monetary incentives, performance targets, and relational control mechanisms such as performance reviews, mentoring sessions, quality circles, and team-building activities.[1] The majority of research on these instruments has focused on their direct effect on employee performance, despite the centrality of training for employee productivity. As a response, scholars have called for more empirical evidence on how these instruments affect performance through employee training (Hoffman & Stanton, 2024; Moers, 2024).

[1]See for instance Mahlendorf and Vogelsang (2024) and Hoffman and Stanton (2024) for recent reviews.

This thesis consists of three studies on the design of management practices for employee development in the context of the information and communication technology (ICT) industry. Specifically, the employees studied in this thesis install, configure, and maintain both household and commercial IT infrastructure. Jobs in this occupational field are characterized by complex and largely non-routine tasks, ranging from technical troubleshooting and hardware installation to customer interaction and on-site support (BIBB, 2023). As a result, these roles require a combination of cognitive and non-cognitive skills, which is highly rewarded in the labor market and increases the costs of acquiring fully trained workers directly from the market (Autor et al., 2003; Deming, 2022; Grundke et al., 2018). At the same time, the ICT industry is characterized by the rapid diffusion of new technologies - such as cloud computing and artificial intelligence - which amplifies the value of adopting new skills and work routines in these jobs (Reljić et al., 2021). These features make the ICT industry a particularly suitable setting for studying how firms can shape employee development using different levers of control. Additionally, the empirical setting of this thesis has valuable properties from an empirical perspective: First, employee performance can be measured objectively along multiple dimensions, including both quantity (e.g. number of completed assignments, average time per assignment) and quality (e.g. customer service ratings, share of returned assignments). Second, most work assignments are completed individually, generating a large number of independent observations and thereby enhancing statistical power.

This thesis is structured as follows. Chapter 1 examines the role of hard information on employees' skills for employee development, performance and job satisfaction. In a field experiment with over 2,500 field service technicians across 125 teams, the managers' access to the results of a biannual skill assessment is randomly removed for 62 managers. Instead, employees and managers are instructed to focus on the employees' suggestions for their own development. The results show that restricting managers' access to this information substantially reduced training investments and overall job performance. Contrary to a large literature in psychology, the findings suggest that the reduction in employee monitoring did not have a positive effect on job satisfaction. Instead, employee satisfaction with the training process, the supervisor, and the overall job appears to be reduced. Analyzing the mechanisms in more detail, the results show that the treatment reduced both employees' and the managers' willingness to invest in employee development. Specifically, employees are less likely to take part in the skill assessment and managers are less likely to schedule development meetings. These results underline the importance of hard information as a commitment device for employee development for managers and employees alike, and showcase the positive side of supervisor monitoring.

In Chapter 2, the role of supervisor interactions in employee learning is studied from a different angle. In this chapter, it is examined how employee performance changes in response to workplace errors. Using performance data from 3,200 employees and 1,100 individual error records from 50 technical service firms, this paper shows that employee performance increases substantially after a manager has confronted an employee with an error reported by an external party. Performance increases gradually until all employees have been contacted by their managers, and remains at a higher level six months after the confrontation and beyond. Heterogeneity analyses suggest that both high- and low-performers initially respond to the intervention, but only the latter group exhibits a sustained improvement in long-term performance. As low-performing employees exhibit flatter learning curves, the manager intervention is more likely to occur when they are still learning. Further analyses show that the intervention by the manager produces negative spillovers in other dimensions of the job, as customer service quality declines after the intervention. This effect is more pronounced for technical errors, which is consistent with the interpretation that the intervention diverts attention toward technical aspects of the job at the cost of other, non-technical aspects. Taken together, these findings further highlight the positive, but also some negative effects of managerial contact on employee development and performance.

Finally, Chapter 3 explores the direct influence of management practices on the integration of new tasks into existing job routines. In a field experiment with over 800 service technicians in 15 firms, half of the workforce receive a temporary bonus for sales activities, which many technicians previously did not consider a core aspect of their job. The results show that the bonus strongly increases employees' engagement in sales activities and downstream product purchases. The results do not indicate that this effect comes at the cost of lower customer satisfaction. Crucially, sales activities not only increase when the incentive was in place, but also three months after the incentive period had ended. Using a post-experimental survey, different channels for habit formation are explored. The results suggest that employees "acquire a taste" for sales activities and customer interaction more generally, as they show increased levels of intrinsic motivation for sales activities, as well as for another customer-oriented task. Overall, the experiment demonstrates how firms can support the formation of new work routines through monetary incentives. Contrary to the literature on the crowding-out effects of monetary incentives, the results suggest that incentives can also potentially "crowd-in" intrinsic motivation and thereby lead to sustained performance and profit increases beyond the incentive period.

Taken together, this thesis shows how firms shape employee development through information, incentives, and managerial interaction. Across all settings, the results show that these instruments have persistent effects on various measures of performance, employee motivation and well-being.

1 Managing Skills in Organizations - Evidence from a Field Experiment

This chapter is based on Grabe and Sliwka (2025)

Abstract. We investigate the value of skill management in organizations through a natural field experiment with 2,582 service technicians. Employees in both treatment and control groups could self-assess skills and propose training measures to managers. However, restricting managers' access to objective skill information reduced training intensity, work performance, and job satisfaction – indicating underinvestment in human capital, even with employee-initiated development. Using detailed personnel records and survey data, we show that the intervention dampened both employees' efforts to identify training needs and managers' engagement in staff development. In particular, high-skilled employees received less training aimed at broadening their skill set, and performance declined due to longer completion times for complex assignments. These findings highlight the pivotal role managers play in fostering skill development and workforce productivity.

Keywords: Skill Management, Training, Performance, Field Experiment
JEL Codes: J24, J28, M12, M53

This paper is co-authored with Prof. Dirk Sliwka from the University of Cologne. I developed the experiment with support from Prof. Sliwka. I was responsible for the implementation of the experiment and the post-experimental survey. I collected the data and conducted the data analysis with support from Prof. Sliwka. I prepared the first version of the draft which was then jointly revised by Prof. Sliwka and me.

1.1 Introduction

Human capital is a critical driver of firm performance, which is why firms have a vested interest in a highly skilled workforce (Bartel, 1989; Black & Lynch, 1996; Bloom et al., 2010). To benefit from technological innovation, firms must ensure that their employees are prepared for the ever-evolving work requirements (Acemoglu & Restrepo, 2022; Autor et al., 2024; Brynjolfsson et al., 2018). As acquiring new skills is central to this transformation, understanding and developing employee skills has become crucial for firm success. Historically, measuring and monitoring employee skills has been challenging and costly. However, the widespread adoption of learning technologies has significantly lowered the barriers to systematic skill management.[1] While there exists a large literature in labor economics on the strong link between employee skills, performance and earnings (Acemoglu & Autor, 2011; Alfonsi et al., 2024; Bapna et al., 2013; Hanushek et al., 2015; Mincer, 1974), and a growing literature on management practices that are conducive to employee development (Buell et al., 2022; Cai, Chen, et al., 2024; Friebel et al., 2022; Sandvik et al., 2020, 2025), the internal mechanisms by which firms invest in employee skills remain under-explored.

We make use of a unique setting which allows us to study a skill management process in detail using a rich set of information on skill assessments, training participation, supervisor interactions as well as productivity data and survey results. In a natural field experiment involving 2,582 field service technicians from a large technology company, we vary managers' ability to monitor and manage employee skills. The skill management process of the firm is as follows: Employees in the organization are regularly invited to take part in online assessments on job-specific technical skills. Managers are instructed to hold quarterly development meetings where they review the results of these assessments and discuss training recommendations from the employee. Based on the assessment and the discussion in the meeting, the manager then decides which training the employee receives and when. In our experiment, managers (and their respective teams) are randomly allocated to a control or treatment group. Employees in both groups were encouraged to recommend training measures to their managers after completing the skill assessment. Prior to the intervention and in the control group, employees were expected to share additional information with their managers, such as the dates when the assessments

[1] Despite the importance of skills in the labor economics literature, there is little consensus on a definition of employee skills. In this paper, we focus on technical skills, which are "combinations of cognitive and non-cognitive skills used to accomplish specific tasks" (Brunello & Wruuck, 2021, p. 1146). Therefore, skill management includes all activities to facilitate employee development and the application of acquired skills in the interest of the firm.

were conducted and the answer to every question in each assessment. Employees in the treatment group were told *not* to report assessment outcomes and managers were instructed *not* to demand them. In other words, by removing managers' access to hard information on employees' skills, their ability to actively manage these skills is reduced while employees retain their autonomy to identify skill gaps and propose training investments. This change allows us to examine how skill management by managers affects the allocation of training and the work performance of employees.

Our key results are the following: Reducing managers' ability to monitor employee skills lowered the overall training intensity substantially as the treatment reduced the number of trainings booked by about 50%. It also led to a decline in employee performance by causing significantly longer task execution times. A back-of-the-envelope calculation indicates that this decline in efficiency would translate into approximately USD 1,040,000 in additional annual personnel costs. Furthermore, contrary to the common view that monitoring diminishes psychological well-being, our results demonstrate that its removal did not improved job satisfaction but, in fact, led to a significant decline.

To explore the underlying mechanisms in more detail, we examine the roles of employees and managers in the skill management process. We begin by documenting a complementarity between the actions of both parties: training intensity is highest when employees assess their skill level to identify training needs, and managers take time to discuss those needs with their staff on a regular basis. Conversely, training intensity is substantially lower when either party does not contribute to the development process.

We then demonstrate that the treatment reduced the contributions from both employees and managers to this process. On the one hand, employees are less likely to participate in the assessment, especially those with a lower skill level. On the other hand, managers reduce their engagement in particular for high-skill employees, which may explain why the reduction in training is largest in this group. Survey evidence further shows that the motivation of high-skilled employees to demonstrate their competence was reduced when assessment scores are no longer revealed to managers. Although a decline in human capital investments among already high-skilled employees may not appear problematic at first glance, we show that the intervention undermined the acquisition of *new* skills among these employees. In line with this observation, we find that productivity losses are driven by longer completion times for complex tasks, which require an advanced skill set. Taken together, our findings underscore the vital role of skill management – not only in mitigating skill gaps among low-skilled employees, but also in sustaining continuous learning and upskilling among the high-skilled.

Our paper contributes to the study of training within firms. While there is a growing literature that quantifies the direct and indirect returns of training in organizations (Adhvaryu et al., 2023; De Grip & Sauermann, 2012; Espinosa & Stanton, 2023), "little work has covered how firms decide on training investments" (Hoffman & Stanton, 2024, p. 78). Employees often take up training on their own initiative, which can be inefficient, as those who are most likely to gain from training are the least likely to take it up (Delfino et al., 2026; Sandvik et al., 2025). Therefore, the standard rationale behind the managerial assignment of training is to focus on low- or under-performing employees with the goal of raising performance up to that of the average employee (Adhvaryu et al., 2022; Diaz et al., 2025). In contrast to this approach, we find that investing more time and effort in broadening the skill set of high-performing employees can be more fruitful in an environment that is characterized by frequent technological change. Our paper further contributes to the research on management practices (Bloom et al., 2007, 2012; Gosnell et al., 2020). Several studies have shown that monitoring employees is positively related to firm-level productivity and profitability[2], as it solves classical moral hazard problems (Holmström, 1979; Merchant & Van der Stede, 2017). On the other hand, the introduction of a monitoring technology is often perceived critically by employees, potentially leading to lower work morale, feelings of distrust and lower performance (Campbell et al., 2011; Christ, 2013; Falk & Kosfeld, 2006; Nebeker & Tatum, 1993). We show that the monitoring of skills can be beneficial for productivity and employees' job satisfaction, because it increases investments in the development process among both managers and employees. Finally, our results also contribute to the literature on career and image concerns in organizations (Bénabou & Tirole, 2006; Holmström, 1999) by providing evidence on employees' motivation to signal their capabilities to their superiors. Prior research shows that in situations of (managerial) observation, people manipulate signals of skill (Burks et al., 2013; Ewers & Zimmermann, 2015) and concentrate their effort on tasks that allow them to showcase their ability (De Janvry et al., 2023; Katok & Siemsen, 2011b, 2011a). While these phenomena have primarily been documented in laboratory settings, field (experimental) evidence on signaling in the workplace is scarce.

The paper proceeds as follows: Section 1.2 illustrates the operating environment of

[2]For instance, Jackson and Schneider (2015) show that moral hazard explains why auto mechanics are reluctant to use checklists to monitor their performance, giving managers in a retail bank access to objective performance measures increases profits (Manthei & Sliwka, 2019), randomly installing GPS tracking devices in a Nigerian trucking firm increases average speed (Rochambeau, 2017), increasing the visibility of the monitoring technology improves task performance in some dimensions (Jensen et al., 2020), and introducing performance reviews in a retail chain raises profits (Manthei et al., 2023).

the firm and their skill management process. Section 1.3 lays out the experimental manipulation and shows descriptive statistics of key dependent variables. In Section 1.4 we show causal effects of the intervention on trainings, work performance, and employee well-being. We then analyze potential mechanisms in Section 1.5 to understand the role of managers and employees in the skill management process. In Section 1.6 we show how these mechanisms affect training and work performance. Section 1.7 concludes.

1.2 Institutional Setting

We run the experiment in the technical service division of a large service organization. The organization employs service technicians operating in 249 teams, 125 of which participate this experiment. Service technicians mostly work independently and their primary responsibility is to install and maintain the firm's products and services. While some assignments can be performed remotely, most of the work is done by one technician at the client site, i.e. in private homes or offices sites. Within each team, every employee is responsible for a designated geographical area. Employees receive their daily work schedule from a regional dispatcher. The team manager coordinates the content of the work schedule with the dispatcher and ensures that the technicians on their team have the skills required to complete their assignments within the allotted time frame (see Figure 1.4 for an overview of the relationship between employee, manager and regional dispatcher). Given the technical nature of the work tasks and ongoing technological changes in the industry, continuous skill management is essential for maintaining high levels of productivity and customer satisfaction. An overview of the firm's skill management process is presented in Figure 1.1. The first step in this process involves employees participating in regular knowledge assessments to evaluate their skills. Assessments are conducted twice per year for each of six core technical skill domains (hereinafter referred to as "skills"). Each employee has a skill profile specifying a subset of the six core skills, and these skill profiles determine the work assignments that can be allocated to them. Most employees have two or three skills and they are automatically invited to the respective assessments. Hence, a typical employee is supposed to take part in around five skill assessments per year. Supervisors may also assign new skills to employees in order to broaden their skill sets. Skill assessments are administered through multiple-choice questions on an online platform, accessible via tablet, phone, or laptop. Employees receive an automated invitation and are given a designated one-hour time window during which no work assignments are scheduled. Each assessment contains 20 questions, randomly selected from a pool of 60 questions per skill. If an employee answers a question incorrectly, the correct answer, along with a brief explanation, is provided before they proceed to the next question. All questions are created by experienced technicians and reviewed by the works council to ensure that they cover the actual work content appropriately.

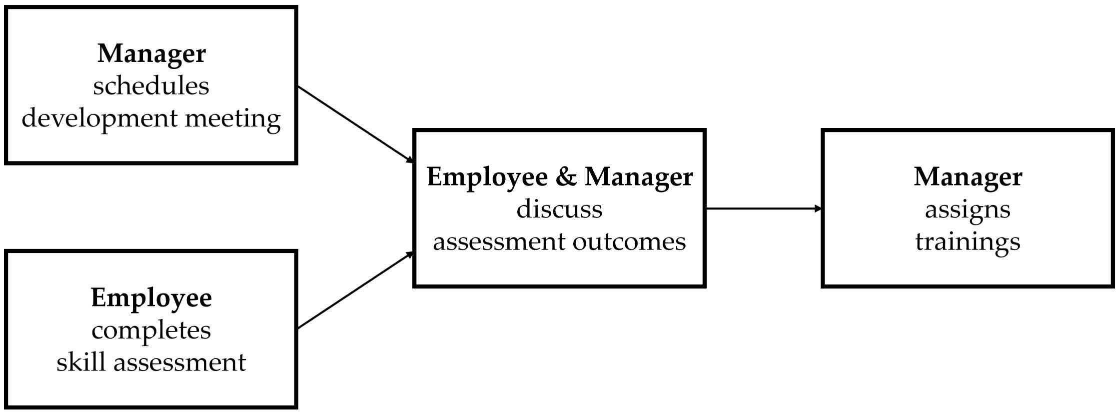

Figure 1.1: Overview of the Skill Management Process of the Firm.

After completing the assessment, employees receive an overall rating of their current skill level, along with a detailed record of their responses to every question. They can then choose from a set of suggested training measures, which they may recommend directly to their manager. During this step, employees can also write a personal message that will be transmitted along with their suggestions (see Figure 1.5 for screenshots). In the following step, managers and employees meet to discuss the results of the skill assessment and the employees' training proposals. Training usually serves one of two purposes: Training should either fill gaps in skills employees already have ("skill preservation") or broaden their skill set by introducing them to a new technology or expose them to more advanced, non-routine problems ("new skill acquisition"). It is thus possible that two employees receive training on the same broader topic, but the content differs because the training has a different purpose. Given the technical nature of the work, training mostly takes place in person. In these training sessions, employees practice on real products with their daily tools under the supervision of an instructor.

Before the intervention and in the control group, employees are asked to bring their evaluation sheets to the development meetings. Managers then decide on the allocation and scheduling of training sessions, based on the employees' proposals and the results of the skill assessment. Since the introduction of the learning platform, the firm's management and its works council have been debating the appropriate level of monitoring within the organization.[3] The works council argues that showing detailed assessment results to managers is unnecessary and potentially degrading, suggesting that employees might be discouraged from participating in assessments due to fear of being blamed for low skills. To gain a deeper understanding of these concerns, we conducted ten short phone interviews with five supervisors and five employees from different teams, who were referred to us by the firm. In these interviews, some employees indeed expressed that fear of being blamed for poor outcomes might contribute to a reluctance to participate in skill assessments or to share their results in the development meeting. Although the firm's leadership has historically been in favor of monitoring employee outcomes, they were open to reducing it if doing so would improve assessment participation and encourage training, particularly among low-skill employees.

1.3 Experimental Setup

We randomly assigned 125 teams (N = 2,582 employees) to either a treatment or a control group. In both groups, employees were encouraged to propose potential training measures to their managers. In the treatment group, employees were explicitly instructed *not* to share their assessment outcomes with their managers, and managers were told not to request them, but rather to focus on the training measures proposed by their employees. The firm is organized into six departments that comprise a total of 37 regions. Treatment assignment was stratified at the regional level. We further included team size, average team assessment participation, and assessment scores as stratification variables. The experiment was conducted from July 1, 2022, to March 31, 2023. Employees in both groups were informed about upcoming changes to the learning platform via email and directly through the platform when they complete a skill assessment (see Figure 1.5). All managers and employees were invited to participate in virtual Q&A sessions, which were split by treatment group and role. The post-experimental survey was distributed via email on behalf of the CEO in January 2023. All participants who completed the survey received compensation of €7 via bank transfer.[4] Unless stated otherwise, we use all

[3]In Germany, employees have a right to set up an employee-elected works council in establishments with more than 5 employees. Firms need the consent of the works council when implementing policies to monitor employees' performance. § 87 (1) of the Works Constitution Act [BetrVG], for instance specifies that "The works council has a right of co-determination [..] in the introduction and use of technical devices designed to monitor the behavior or performance of the employees".

[4]This particular survey was conducted by the researchers and respondents were informed that the company would only receive aggregated responses and anonymity is preserved. Compensation for survey

Table 1.1: Balance Table

	Control (N=63)		Treatment (N=62)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Assessment Participation	0.81	0.16	0.80	0.17	-0.02	0.54
Assessment Score	0.00	0.28	0.00	0.34	0.00	0.95
Number of Trainings	0.25	0.30	0.24	0.30	-0.01	0.79
Supervisor Meetings	0.86	0.34	0.78	0.41	-0.07	0.35
Troubleshooting (in min.)	34.41	3.79	34.05	4.14	-0.36	0.62
Customer Service (in min.)	7.19	1.27	6.68	0.98	-0.51	0.02
Quality Control (in min.)	10.34	2.08	10.54	2.28	0.20	0.63
Documentation (in min.)	4.87	0.36	5.02	0.48	0.15	0.06
Team Size	20.00	2.16	20.13	2.80	0.13	0.77

Note: Table shows pre-treatment means and standard deviations from all teams included in the experiment. Assessment Participation denotes the average participation for the assessments every member of the team has been assigned to. Assessment Score indicates the average standardized ratio of correctly answered questions per assignment. Number of Trainings denotes the average number of trainings per employee in the last 6 months. Supervisor Meetings is a team average of the number of development meetings per employee. Troubleshooting, Customer Service, Quality Control and Documentation denote the weekly average completion time for each task in minutes.

available data from the beginning of 2022 until the end of the post-experimental survey in March 2023. For the data analysis, we excluded employees who changed teams before the end of March 2023, as they may have been part of different treatment conditions. Table 1.1 provides summary statistics for relevant pre-experimental outcomes for the treatment and control group. In addition to the stratification variables – assessment participation, standardized assessment score, and team size – we also consider key dependent variables for the analysis, such as the number of training sessions per capita and the average duration of key work tasks per team.[5] Prior to the intervention, employees on average took part in about 80% of the skill assessments relevant to them and attended about 0.5 training sessions per year.

participation was paid from university research funding.

[5]Note that it was not possible to measure the number of trainings and development meetings on an individual level prior to the experiment. Instead, data collection for this time period was limited to the team level.

1.4 Results

1.4.1 Trainings

First, we examine the effect of the treatment on employee development, as measured by the number of trainings assigned by the supervisor. The average treatment effect on the number of trainings is depicted in Figure 1.2 and the corresponding regression results are displayed in Table 1.2, where we also analyze the likelihood of an employee receiving at least one training in the treatment period. In columns (2) and (4), we control for pre-treatment participation in the skill assessment and the number of skills prior to the experiment. As Figure 1.2 shows, the number of trainings per employee in the first 6 months of the treatment period is reduced by 0.19 compared to the control-group mean of 0.37, which corresponds to a 51.89% reduction ($p = 0.036$). Estimating the extensive margin effect – i.e. the likelihood of receiving any training at all – we find that employees who are not allowed to bring their assessment results to the development meeting are 7.23 percentage points less likely to receive any training in the first 6 months of the treatment period, which translates to a 46.85% reduction ($p = 0.03$) in comparison to the control group (see columns 3 and 4 of Table 1.2). Therefore, reducing managers' ability to directly monitor and manage employee skills substantially reduces training intensity and the overall likelihood of receiving training.[6]

[6]In Table 1.10, we replicate our results using a Pseudo-Poisson ML estimator for the treatment effects in columns (1) and (2) of Table 1.2 and a logit model to estimate the treatment effects depicted in columns (3) and (4). We further re-estimate the treatment effect on the number of trainings on the team-level where we can control for pre-treatment values and target a longer time period (12 months instead of 6), which yields similar results.

Table 1.2: Treatment Effects on Trainings

	Num. of Trainings		Prob. of Training	
	(1)	(2)	(3)	(4)
Treatment	−0.194**	−0.213**	−0.072**	−0.077**
	(0.091)	(0.097)	(0.033)	(0.035)
Prior participation		0.268***		0.109***
		(0.070)		(0.025)
Number of Skills		0.021		0.012
		(0.031)		(0.010)
Constant	0.374***	0.121	0.154***	0.042
	(0.077)	(0.119)	(0.027)	(0.044)
Observations	2582	2234	2582	2234

Note: Table shows estimates from linear regressions on the number of assigned trainings per employee and the likelihood of receiving at least one training assignment. Columns (1) and (2) use the number of trainings as a dependent variable, which is the sum of all trainings associated with key skills. In columns (3) and (4) the dependent variable is a binary indicator for receiving training which takes value 1 if an employee has received at least one training and 0 otherwise. Prior participation captures the participation rate in the skill assessment prior to the experiment. Standard errors are clustered at the team level and reported in parentheses. **, *** indicate significance at the 5% and 1% levels, respectively.

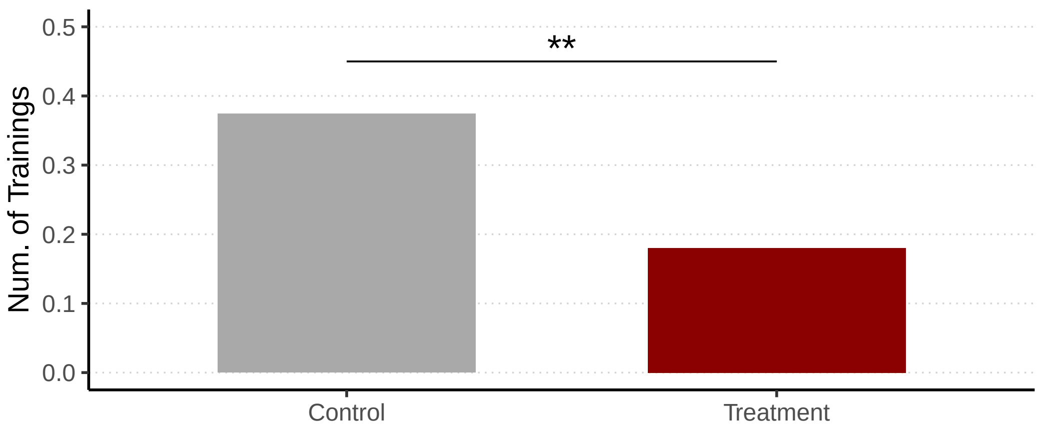

Figure 1.2: Treatment Effect on the Number of Trainings per Employee

1.4.2 Work Performance

To understand the effects of the intervention on employee performance, we utilize the firm's performance tracking system, which measures the time a technician takes to complete a task in a given work assignment. We focus on the four work phases that are common for almost every work assignment and account for 95% of all observations in the data set.[7] The variable Customer Service (M = 6.27 min) includes all tasks associated with customer interaction prior and during the appointment. For most clients, employees are required to call the customer to understand their infrastructure and their current issue. This short conversation often serves as a precursor to on-site troubleshooting. Troubleshooting (M = 43.29 min) is the core of a technician's job and involves all tasks associated with problem-solving at the client site. Troubleshooting is the most time-intensive task where technical knowledge and training are most likely to play a significant role. Quality control (M = 10.22 min) follows troubleshooting and involves assessing the client's infrastructure, typically after resolving the issue. If any further problems remain, this serves as documentation for a follow-up visit. Finally, documentation involves all tasks associated with paperwork or data entry (M = 4.48 min). As there is a slight imbalance in prior task times (see Table 1.1), we perform difference-in-differences regressions for the average time needed to complete each step in the work assignment. We include team fixed effects and assignment-specific time trends to account for seasonal changes in product sales (e.g. holiday seasons).[8] As Table 1.3 shows, reducing the access to employee skill assessments reduces employee performance, as is evident from the increased average assignment times in two key phases. Column (1) shows that teams in the treatment group need 0.51 minutes longer to complete the same customer service assignments as teams in the control group. In comparison to the control group mean prior to experiment, the treatment thus caused an 8.11% increase in the average time spent on customer service tasks (p = 0.046). The average time spent on troubleshooting activities increases by 1.38 minutes (see column 2) or 3.19% (p = 0.074). We do not find effects on the last two steps of the work assignment, such as quality control (column 3) and documentation (column 4). To rule out that the additional time spent on troubleshooting and customer interaction had a positive effect on work quality, we also study the effect of the treatment on assignment quality. As shown in Table 1.11, we find no evidence

[7] One row in the data set represents the average time to complete a specific activity in a given week across all members of one team. The tasks presented in Table 1.3 cover 95% of all rows in the data set. The remaining tasks concern, for instance, the delivery of equipment to a work site or manual programming.

[8] A detailed analysis regarding the nature of the work assignment follows in Section 1.6, where we further differentiate work assignments by their degree of complexity.

Table 1.3: Treatment Effects on the Average Time per Task

	Customer Service	Troubleshooting	Quality Control	Documentation
	(1)	(2)	(3)	(4)
Treatment	0.509**	1.381*	−0.051	−0.033
	(0.252)	(0.766)	(0.462)	(0.100)
Observations	33 880	42 843	34 734	38 191
Team FE	Yes	Yes	Yes	Yes
Type × Week FE	Yes	Yes	Yes	Yes
DV mean	6.27	43.29	10.22	4.48

Note: Table shows average treatment effects from difference-in-difference regressions on the average time per task in minutes. All models include team as well as Type × Week fixed effects. Assignment types are usually associated with a specific technology, but do vary with respect to their degree of complexity. Pre-experimental averages from the control group are depicted in the last row of the table. Standard errors are clustered at the team level and reported in parentheses. *, **, indicate significance at the 10% and 5% levels, respectively.

that the treatment had an effect on customer satisfaction or the share of assignments that required a follow-up visit. A back-of-the-envelope calculation illustrates that the personnel costs associated with the treatment can be substantial. Even if employees complete only two in-person customer visits per day, the estimated time increase of the treatment is equal to 3.35×10^4 hours per year. This corresponds to the annual working hours of 20 full-time employees or about 1 million USD in estimated personnel costs.[9]

1.4.3 Employee Perceptions

While most research in economics documents positive effects of monitoring on performance, psychological research suggests that increased levels of monitoring can reduce employee well-being (Deci et al., 1989; Nebeker & Tatum, 1993; Ravid et al., 2023) – a point often brought up by the popular press and in our discussions with the firm (Kantor & Sundaram, 2022; Murty & Karanth, 2022; Shrikant, 2023). Therefore, one might expect that removing managers' ability to observe the outcomes of the skill assessment would reduce the perceived pressure and lead to higher job satisfaction. To study the effect of the treatment on job satisfaction, we use data from the post-experimental survey, which can be linked with individual skill assessment and training data (see Section 1.8.3 for all

[9] In this calculation, we neither consider the time saved by the reduction in training, nor a potential relative decrease in wages that stems from the reduction in skill acquisition. Therefore, our calculations should be considered as an upper bound. Wage information was retrieved from the German Union for Technical Workers (IG Metall).

Table 1.4: Treatment Effects on Employee Perceptions

	Job Satisfaction	Supervisor Support	Turnover Intention
	(1)	(2)	(3)
Treatment	−0.200**	−0.214**	0.088
	(0.077)	(0.097)	(0.079)
Constant	0.094*	0.098	−0.040
	(0.053)	(0.062)	(0.055)
Observations	775	848	819

Note: Table shows estimates from linear regressions on standardized measures of job satisfaction, supervisor support and turnover intentions as reported by the employees in the post-experimental survey. Job satisfaction and turnover intentions are measured using single-item measures from the Linked Personnel Panel (Kampkoetter et al., 2016), our measure of supervisor support is based on a four-item survey from Decius, Schaper, and Seifert (2021) and a three-item scale from Udris and Riemann (1980). All questions are answered on a 7-point Likert scale from 1 = 'strongly disagree' to 7 = 'strongly agree'. Standard errors are clustered at the team level and reported in parentheses. *, ** indicate significance at the 10% and 5% levels, respectively.

survey items). The overall participation in the post-experimental survey is 36.76%. While survey participation was larger in the control group, participants in both groups exhibit very similar characteristics with respect to prior participation in skill assessments and achieved assessment scores (for more details see Table 1.16). Analyzing treatment effects on psychological outcomes shows a universally negative impact: As shown in Table 1.4, the treatment significantly reduced job satisfaction (column 1, survey item "*All things considered, I am quite satisfied with my current job.*") and supervisor support (column 2), which refers to the perceived attention and effort supervisors invest in their employees' development (seven items, e.g. "*My supervisor is interested in what I'm currently learning.*"). It is important to recall that in the control group, employees had the same access to the skill assessments and also were encouraged to share their training needs with their managers. Nevertheless, the active management of skills promoted job satisfaction, a finding we explore in more detail below.

1.5 Mechanisms

Having established that reducing managers' ability to actively manage employee skills reduces training intensity, performance, and job satisfaction, we now investigate the underlying mechanisms in more detail. As illustrated in Figure 1.1, successful skill management requires both the employee and their respective supervisor to invest time and effort: Employees need to be willing to identify skill gaps by taking part in the skill assessment, and managers need to schedule development meetings to discuss assessment results and assign training. Indeed, a simple descriptive regression of training intensity on test participation, meeting incidence, as well as their interaction shows a complementary relationship between both contributions (see column 2 in Table 1.12). That is, training intensity is low when either of the two contributions is absent. In the following, we study the contributions of both sides in turn. We first consider the employee perspective, where we analyze treatment effects on participation and results in the skill assessments. Then, we study manager behavior where we investigate the role of managerial attention in the skill management process.

1.5.1 Employee Behavior

To study treatment effects on employee actions, we focus on participation in the skill assessment and the outcomes of those who participated.[10] Employees are only invited to a skill assessment if they have been assigned to the corresponding skill. The key outcome variable that we consider here is thus average participation for assessments associated with these three skills. Prior to the experiment, the average share of correctly answered questions was 53.03% in the most widespread skill assessment. In the second and third most widespread skill assessments, this share was 64.02%, and 71.64%, respectively. Our final data set includes 19,447 assessments results, which we standardize for each assessment separately and then create the mean standardized score across all three assessments. Table 1.5 shows treatment effects on assessment participation and the average standardized score. As column (1) shows, the treatment reduced the average

[10]We focus on the three most important skill assessments that correspond to the majority of work assignments of the firm and for which both pre- and post-experimental data are available. Figure 1.6 shows that the three basic skills cover 89% of all work assignments. In Table 1.13, we show treatment effects for all skill assessments in the treatment period across all skills. While it is possible to re-take the assessment, we find only six instances in our data set where an employee participates twice in the same assessment. Therefore, we only consider assessment scores in the first try for every assessment.

participation rate by 6.59 percentage points from 79.03% to 72.43% ($p = 0.004$).[11] Hence, rather than encouraging employees to take part in the assessment to learn about their own skill gaps when they can do so in private, the intervention reduced the likelihood that employees assess their own skills. We find no evidence, however, that the treatment affected the average assessment result of those who took part in the skill assessment.

A key conjecture at the outset of the experiment was that reducing managers' ability to observe assessment outcomes affects employees differently depending on their prior skill level. If employees with lower skill levels are discouraged from performing assessments due to fear of negative judgement, the treatment may have positively impacted participation by ensuring results remain private.[12] To investigate this, we classify those employees that had taken part in at least one of the focal skill assessments prior to the intervention and achieved an above-median score as *high-skill,* and the remainder as *low-skill.*[13] The corresponding results are shown in the columns "high-skill" and "low-skill" of Table 1.5. Contrary to our expectations, column (2) shows that the reduction in assessment participation seems to be driven by low-skill employees ($p < 0.001$). Hence, the intervention did *not* encourage this group to take part in the assessment. On the other hand, participation among high-skill employees remains unaffected ($p = 0.109$). If anything, we find weak evidence that the treatment reduced assessment scores of high-skill employees relative to those with lower skills ($p = 0.071$).[14]

Hence, rather than encouraging employees with lower skill levels to assess their own skills by reducing the fear of being blamed for bad outcomes, the treatment even reduced assessment participation particularly in this group. Assessment scores are somewhat reduced for the high-skill employees – likely because they can no longer impress their

[11]In Table 1.19, we analyze whether the treatment has reduced the employees' reasoning about the priority of the skill assessment. We find no evidence that the treatment has lowered the perceived priority of the assessment. Therefore, employees are still aware that they are expected to complete the skill assessment, but rather consciously decided to avoid it now that non-participation is not observable for the manager.

[12]Our pre-registered hypotheses here were that the treatment (i) reduces assessment participation and performance of highly skilled employees but (ii) increases participation for employees with lower skills.

[13]Note that prior non-participants are thus also classified as *low-skill*. In Table 1.14, we show that there is a strong correlation between test participation and test outcomes, i.e., past non-participants have substantially lower test outcomes when taking part in a test now and current test participation is higher for employees with better past test outcomes. Additionally, Figure 1.8 shows that prior non-participants on average have fewer skills compared to any other group and it is therefore unlikely that these prior non-participants are in fact high-skilled employees who were not motivated enough to participate in the skill assessments.

[14]In Table 2.5, we re-run the analyses with an interaction term instead of a sample split, which yields the same results. We also look at more specific subgroups comparing test participation and test scores between treatment and control group for the prior non-participants and by quartile of prior assessment scores. The results are displayed in Figure 1.7 in the Appendix. Notably, the reduction in test participation is driven by those employees who had not taken part in prior tests, while high-performing employees seem - at least to certain extend - reduce their assessment scores in response to the treatment.

Table 1.5: Treatment Effects on Assessment Participation and Assessment Scores

	Assessment Participation			Assessment Scores		
	Full Sample	High-Skill	Low-Skill	Full Sample	High-Skill	Low-Skill
Treatment	−0.066***	0.007	−0.093***	−0.015	−0.090	0.044
	(0.022)	(0.023)	(0.028)	(0.050)	(0.056)	(0.063)
Constant	0.790***	0.819***	0.761***	−0.054	0.397***	−0.440***
	(0.015)	(0.017)	(0.018)	(0.034)	(0.039)	(0.042)
Observations	2321	1042	1190	2210	979	1009

Note: Table shows estimates from linear regressions of assessment participation (columns 1-3) and assessment scores (columns 4-6) on a treatment indicator. Participation denotes the number of assignments an employee has taken part in during the treatment period, divided by the number of assignments they have been assigned to. Assessment scores depict the average standardized ratio of correctly answered questions on the knowledge assessment. Columns 1 and 4 show results for the full sample. For the sub-sample analyses, we split the sample based on the variable 'High-Skill' which takes a value of 1 for all employees with an above median average assessment score prior to the experiment (columns 'High-Skill') and value 0 otherwise (columns 'Low-Skill'). Standard errors are clustered at the team level and shown in parentheses. *** indicates significance at the 1% level, respectively.

supervisors through good test scores. It has often been claimed that reputational concerns are key drivers of effort in organizations, or as put by Holmström (2017) in his Nobel Prize lecture: "the craving for appreciation and the desire to impress superiors" (p.1772). To study the role of such reputational concerns in more detail, we turn to the post-experimental survey where we included two items to assess the role of reputational concerns for taking part in the skill assessment ("*I participate to demonstrate to others that I am competent*.") or avoiding it ("*If I don't participate, it's because I expect to perform poorly.*"). First of all, the motivation to demonstrate competence is substantially stronger than the fear of being viewed as incompetent. On a scale from 1 ("Fully disagree") to 7 ("Fully agree") the mean is 4.71 for the former and 2.19 for the latter. To see how the treatment affects these motives, we re-run the analyses above with the responses to each item as the dependent variables. As column (1) in Table 1.6 shows, participants in the treatment group are significantly less motivated to participate in the knowledge assessment to demonstrate their ability to their supervisor. Column (2) shows that this effect is indeed driven by high-performing employees. In contrast to our prior expectations but in line with our observations regarding the treatment effect on assessment participation, the treatment does not reduce the (already low) stated motive to avoid the skill assessment in fear of being exposed as incompetent (column 4 in Table 1.6). While low-skill employees indeed state significantly more often that they avoid the assessment due to being afraid to do poorly than high-skill employees (columns 3 and 4), there is no evidence for differential

Table 1.6: Signaling and Reputational Concerns

	Demonstrate Competence			Fear of Incompetence		
	Full Sample	High-Skill	Low-Skill	Full Sample	High-Skill	Low-Skill
Treatment	−0.196**	−0.375***	−0.004	0.066	0.070	0.094
	(0.078)	(0.104)	(0.110)	(0.069)	(0.102)	(0.110)
Constant	0.090**	0.174***	0.002	−0.030	−0.033	−0.043
	(0.045)	(0.061)	(0.074)	(0.042)	(0.068)	(0.067)
Observations	850	421	364	847	418	364

Note: Table shows estimates from linear regressions on the strength of an employee's competence motif when considering to participate in the skill assessment (columns 1-3) and the strength of an employee's avoidance motif when considering not to participate in the skill assessment (columns 4-6). All dependent variables have been standardized. Questions are answered on a 7-point likert scale from 1 = 'strongly disagree' to 7 = 'strongly agree'. For the sub-sample analyses, we split the sample based on the variable 'High-Skill' which takes value 1 for all employees with an above median average assessment score prior to the experiment (columns 'High-Skill') and value 0 otherwise (columns 'Low-Skill'). Standard errors are clustered at the team level and shown in parentheses. ***, ** indicates significance at the 1% and 5% levels, respectively.

treatment effect on this motive.

We further explored heterogeneous treatment effects on job satisfaction (see Table 1.17). In line with the notion of losing a source of recognition, job satisfaction is reduced in particular among the high-skilled. The treatment effect is weaker (and statistically insignificant) among for low-skill employees.[15]

1.5.2 Supervisor Behavior

In this section, we study the role of managers in the skill management process. To assess the treatment effect on managerial attention, we first investigate how the treatment affected the likelihood that managers conducted developmental meetings and at which frequency. We begin by regressing the meeting incidence (as reported by employees in the post-experimental survey) on a treatment indicator. As the results reported in Table 1.7 show, the treatment reduced the likelihood that at least one meeting is conducted from 0.77 to 0.67, i.e., by about 13.63% (see column 1). Hence, the intervention not only reduces employees' willingness to assess their own skill gaps but also managers' own time investments in employee development. To understand the behavioral drivers of this effect, it is informative to note that the treatment primarily affects the *likelihood* of

[15] The latter may indicate that there are indeed countervailing effects as monitoring may still have led to embarrassment after bad assessment outcomes while its removal also took away a source of constructive feedback and help by supervisors.

conducting at least one meeting. As documented in column 4, the effect on the number of meetings has a similar magnitude which shows that it is the incidence rather than the frequency that differs. In other words, supervisors do not substantially reduce the meeting frequency *after* having held at least one meeting without having access to assessment outcomes.[16] This suggests that managers are *not* reducing the frequency of meetings because they have had worse experiences as a result of the treatment. Instead, it appears that supervisors ex-ante perceive the meetings as being less valuable when they know that they no longer have access to information on employee skills. This further underlines the importance of hard information on skill levels, not only for employees, but also for managers.

In a next step we study how the reduction in meeting incidence affected employees of different prior skill levels. In the regressions reported in columns 2 and 3 of Table 1.7, we regress meeting incidence and the number of meetings on the treatment for high- and low-skill employees separately. Interestingly, we find that the likelihood of having at least one development meeting is, in particular, substantially reduced among high-skill employees by 18.48% ($p = 0.012$). For low-skill employees, the negative treatment effect is less pronounced at 6.87% and not statistically significant.[17] Apparently, as high-skill employees no longer can demonstrate their abilities (as shown in Table 1.6) the managers also find it less rewarding to meet with them, for instance as it is harder to provide tangible recognition and have a fruitful discussion on their skill set.[18]

The reduction in meeting frequency for high-skill employees (who achieved strong scores in the skill assessments at the outset) may appear less concerning at first glance. However, it is important to recall that as laid out in Section 1.2, a key aim of the firm is not only the preserve the skills for routine tasks, but also to broaden the skill set of employees to adapt to technological changes and foster the acquisition of new knowledge. Indeed employees classified as high-skilled tend to have a larger number of skills that are assigned to them (see Figure 1.8). The substantial reduction in meetings thus may also

[16]To corroborate this we also can study the probability of having more than one meeting conditional on having had at least one. This is indeed essentially unaffected by the treatment: In the sample of employees who had at least one meeting, the estimate of the treatment coefficient is close to zero.

[17]This further shows that individual participation in the skill assessment does not appear to influence the likelihood of arranging development meetings, as the reduction in assessment participation is driven by *low-skill* employees. If assessment participation and managerial attention were directly correlated, the results shown in Table 1.7 would show the opposite effect. This is in line with our interpretation that managers make decisions on the allocation of managerial attention directly in response to the treatment.

[18]Note that the patterns reported in Table 1.6 are not driven by the reduction in meetings for high performers. The results remain virtually unchanged when controlling for meeting incidence or frequency as shown in Table 1.22.

be problematic in this group. We will investigate this in more detail in the next section.

Table 1.7: Treatment Effects on Meetings

	Prob. of meetings			Num. of meetings		
	Full Sample	High-Skill	Low-Skill	Full Sample	High-Skill	Low-Skill
Treatment	−0.106**	−0.144**	−0.052	−0.148*	−0.083	−0.156
	(0.043)	(0.057)	(0.052)	(0.084)	(0.115)	(0.105)
Constant	0.774***	0.778***	0.761***	1.189***	1.160***	1.198***
	(0.025)	(0.034)	(0.032)	(0.054)	(0.068)	(0.077)
Observations	846	419	362	846	419	362

Note: Table shows estimates from linear regressions of the likelihood or frequency of development meetings on a treatment indicator. The dependent variables are a binary indicator that takes value 1 if an employee has indicated to have had at least one meeting with their supervisor (columns 1-3) and the number of development meetings reported by the employee (columns 4-6). For the sub-sample analysis, we split the sample based on the variable 'High-Skill' which takes value 1 for all employees with an above median average assessment score prior to the experiment (column 'High-Skill') and value 0 otherwise (column 'Low-Skill'). Standard errors are clustered at the team level and shown in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 1.8: Treatment Effects by Employee Type

	Num. of Trainings		
	Full Sample	High-Skill	Low-Skill
Treatment	−0.194**	−0.272**	−0.171*
	(0.091)	(0.116)	(0.092)
Constant	0.374***	0.463***	0.348***
	(0.077)	(0.096)	(0.077)
Observations	2582	1043	1191

Note: Table shows estimates from linear regressions on the number of trainings booked by prior skill-level. We split the sample based on the variable 'High-Skill' which takes value 1 for all employees with an above median average assessment score prior to the experiment (column 'High-Skill') and value 0 otherwise (column 'Low-Skill'). Standard errors are clustered at the team level and shown in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

1.6 Training Assignment and Work Performance

In the previous sections we presented evidence showing that the treatment reduced skill assessment participation among low-skill employees, but reduced managerial attention in particular for the high-skilled. In the next step, we explore to what extent this heterogeneity influences the training intensity and type of training received by employees in these two groups, and how this, in turn, can help in understanding the performance effects of skill management. As Table 1.8 shows, the treatment reduced the training intensity in both groups, however the reduction appears to be pronounced among high-skill. This naturally raises the question of whether the reduction in this group is concerning, given that skill levels are already high at the outset. To answer this question, it is important to consider the different purposes of training. As laid out in Section 1.2, the firm distinguishes between training to preserve existing skills and training to acquire new skills. When the reduction in training for high-skill employees is concentrated on the former, it may not be very detrimental or could even increase efficiency by eliminating useless time expenditures. If, however, the treatment affects the latter form of training, it may undercut the firm's ability to adapt and prepare their employees for changing skill requirements.

When managers register an employee for a training, they have to indicate whether the purpose of the training is to maintain an existing skill (referred to as *skill preservation*) or to develop a new skill (*new skill acquisition*).[19] We can make use of this distinction to investigate which types of training is particularly affected in which employee group.[20] Figure 1.3 shows the average number of trainings, split by training type (skill preservation, new skill acquisition) and employee skill level (high, low). Table 1.24 shows the corresponding regression results. While training intensity is lower across both groups and training types, the effect is particularly pronounced for training dedicated to the acquisition of new skills among high-skill employees ($p = 0.01$, panel C) and (to a lesser extent) for skill preservation training for the low-skilled ($p = 0.066$, panel B).

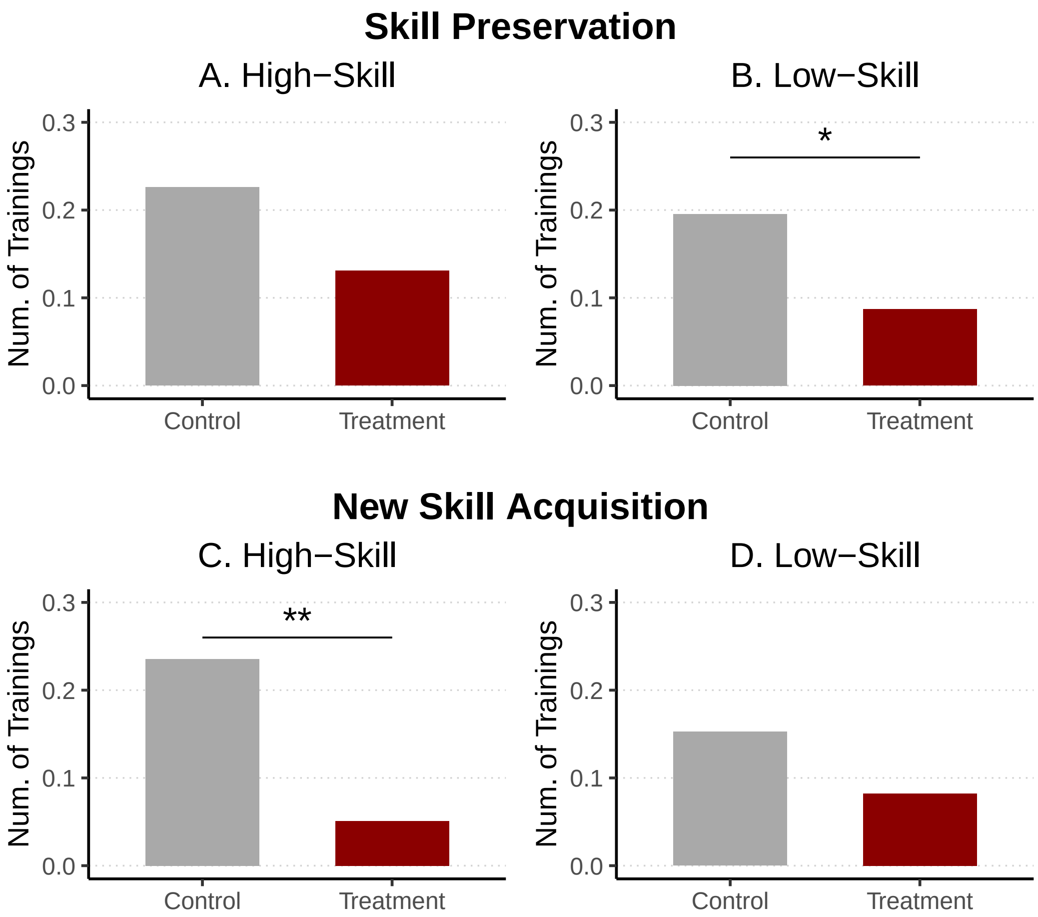

Figure 1.3: Treatment Effects by Employee Type and Training Purpose

[19] Supervisors can also select the option 'not relevant to a specific skill'. However, this option is rarely selected, so we drop this alternative from the analysis.

[20] In Table 1.23, we investigate whether the overall number of trainings dedicated to skill preservation and acquisition is differentially affected by the treatment. We do not find such an effect. If anything, the treatment effect is slightly larger for training dedicated to the acquisition of new skills.

The observation that the reduction in trainings is particularly pronounced for new skill acquisition of high-skill employees may suggest that the overall reduction in work performance is concentrated on work assignments that require a higher level of skill or a broader skill set in general. To investigate this conjecture, we study heterogeneous treatment effects by the degree of assignment complexity. According to the firm's partition of work assignments into 13 different types, each type can be classified as either *routine* or *complex*. Routine assignments often include solving problems that occur on a regular basis or products that require little technical sophistication to work with. Complex assignments often require non-standard custom solutions that have a higher risk of failure. These assignments are either associated with recent technologies that are inherently complex or a sophisticated problem that other technicians were previously unable to resolve. Therefore, assignments can be re-classified from routine to complex. In line with the firm's description, our final classification reveals that routine assignments are performed more regularly and take relatively little time (M = 14.5 min.), while complex assignments take more than twice as long (M = 32.8 min.), and occur less frequently (see Figure 1.9). In Table 1.9, we run separate difference-in-difference regressions for tasks associated with either routine or complex work assignments with the average execution times as the dependent variable. As column (1) shows, there is no discernible treatment effect on the weekly average completion time for routine assignments. However, we indeed find a 3.169 minute increase in the weekly average completion time for complex tasks (column 2), which corresponds to a 9.7% increase relative to the post-treatment average in the control group. In summary, we find that the reduction in training intensity is particularly strong for new skill acquisition among high-skill employees such that the reduction in active skill management in particular leads to a decline in upskilling among these employees. In line with this pattern, the reduction in work performance is concentrated in complex work assignments.

Table 1.9: Treatment Effects on Work Performance by Assignment Type

	Routine	Complex
Treatment	-0.044	3.169^{**}
	(0.184)	(1.568)
Team FE	Yes	Yes
Task FE	Yes	Yes
Product × Week FE	Yes	Yes
DV mean	14.53	32.84
Observations	128 146	21 266

Note: Table shows average treatment effects from difference-in-difference regressions on the average time per task in minutes. Column (1) shows results for work assignments classified as routine assignments, whereas column (2) shows results for complex assignments. For all regressions, we focus on the four key tasks used in the previous analysis. All models include Team, Task, as well as Product × Week fixed effects. Averages from the control group are depicted below the estimates. Standard errors are clustered at the team-level and shown in parentheses. ** indicates significance at the 5% level.

1.7 Conclusion

We studied the value of skill management in organizations through a natural field experiment in a large organization. Our findings highlight that active skill management is crucial for human capital investments and firm performance. Even a slight reduction in managers' ability to monitor and manage employee skills significantly decreased training intensity and productivity. Additionally, contrary to existing literature in psychology and behavioral economics, we observed that reduced supervisor monitoring also led to a decline in job satisfaction.

We further studied the underlying behavioral mechanisms and the roles of employees and managers in this process. Reducing managers' ability to monitor employee skills diminished employees' willingness to assess their own skill gaps, particularly among low-skill employees. Instead of encouraging these employees to identify training needs without fear of blame, the removal of monitoring had the opposite effect. This result highlights the significance of moral hazard issues in training investments. Despite the benefits of human capital gains for employees, there is thus a tendency to under-invest and active skill management helps to mitigate this problem. Moreover, we find that the removal of access to information on employee skill levels also reduced the likelihood that managers initiate development meetings with their employees. Granting supervisors access to hard information about employee skills thus (i) commits employees to invest in their own learning and (ii) fosters managers' own time investments in employee development.

Having documented the main effects and general mechanics of skill management, we set out to understand the connection between the assignment of training and team productivity in more detail. In particular, we study the effects of the interventions for employees of different skill levels. Sandvik et al. (2025), for instance, found that uptake of voluntary training is lower among low-performers, even though they benefit from training the most. In environments undergoing technological changes it is, however, less clear whether prioritizing the skill gaps of low-skill employees should take precedence over broadening the skill set of high-skill employees, as firms constantly need to adapt to changing skill requirements. Our analysis indeed shows that less extensive skill management induced by the treatment particularly reduced managerial attention and subsequent training intensity also among high-skill employees. Moreover, this reduction in training intensity predominantly affected training intended to acquire new skills. In turn, the intervention led to a reduction in work performance in complex, non-routine tasks. These findings add to the literature on heterogeneity in training goals across industries (Alfonsi et al., 2020; Caicedo et al., 2022). In workplace environments with heavily standardized workflows, closing skill gaps among low performers should take priority, as illustrated by a recent study from Adhvaryu et al. (2024). Our results show that for jobs subject to ongoing technological change and innovation crucially benefit from further developing the skill set of high-performing employees.

In summary, our results demonstrate that active skill management and close managerial scrutiny of skill levels are crucial for organizational performance. It fosters the willingness of low-skill employees to assess their own skills, encourages high-skill employees to improve and acquire new skills, and ensures that leaders invest time in helping both groups strengthen their human capital.

1.8 Appendix

1.8.1 Tables

Given that the number of trainings an employee receives in a given time frame can be modeled as count data, we consider alternative specifications for the main analyses conducted in Table 1.2. For the analysis conducted in Table 1.10, we consider a Pseudo-Poisson-ML estimator, which performs well on over-dispersed count data (Silva & Tenreyro, 2006). Estimations are performed in R using the *fixest* package (Berge, 2018). To interpret the results from columns (1) and (2) of Table 1.10, we calculate the incidence ratio by exponentiating the treatment coefficients. Overall, we find that the results are similar to those reported in the main text. For the baseline treatment comparison without covariates, the incidence ratio is 0.481, which is a 48.11% reduction in the number of trainings. Including prior participation in and scores of the skill assessment changes the incidence ratio to 0.479, which is equal to a 47.92% reduction in the number of trainings. In the next step, we re-analyze the treatment effect on whether or not an employee receives any training at all. Again, we find that employees in the treatment group are about 48.96% less likely to receive at least one training during the treatment period. Finally, columns (5) and (6) show team level results, without (column 5) and including pre-treatment data (column 6). While we focus our analyses on the key skills almost all employees possess and for which we have pre and post treatment data, we further analyzed treatment effects on the remaining skill assessments that took place in the treatment period. As Table 1.13 shows, the reduction in assessment participation is strongest among the most widespread skills. In this table, the number of observations denotes all employees that were assigned to the knowledge assessment at the time it took place. As outlined previously, the number of observations is much smaller for less widespread skills. However, including all skills into the assessment does not change our results (see column 7).

Table 1.10: Treatment Effects on Trainings (Robustness Checks)

	Num. of Trainings		Prob. of Training		Num. of Trainings	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	−0.732**	−0.736**	−0.714**	−0.690**	−0.636*	−0.631*
	(0.340)	(0.355)	(0.328)	(0.336)	(0.325)	(0.323)
Prior participation		1.196**		1.413***		
		(0.533)		(0.505)		
Prior trainings						0.036*
						(0.018)
Prior score		0.066		0.103		
		(0.059)		(0.075)		
Constant	−0.983***	−1.963***	−1.702***	−2.873***	2.158***	1.949***
	(0.207)	(0.526)	(0.205)	(0.516)	(0.199)	(0.241)
Level of Analysis	Individual	Individual	Individual	Individual	Team	Team
Observations	2582	2078	2582	2078	125	125
Model	Poisson	Poisson	Logit	Logit	Poisson	Poisson

Note: Table shows estimates from pseudo-poisson regressions on the number of trainings as a dependent variable, which is the sum of all trainings associated with key skills (column (1), (2), (5) and (6). In columns (3) and (4) the dependent variable is a binary indicator for receiving training or not that takes value 1 if an employee has received at least one training and value 0 otherwise. In columns (5) and (6), we use the number of trainings on the team level as the dependent variable. Prior participation captures the participation rate prior to the experiment. Prior score denotes the average standardized ratio of correctly answered questions on a skill assessment. Prior trainings denotes the number of trainings per team prior to the experiment. Robust standard errors (clustered at the team level in columns 1-4) are shown in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 1.11: Treatment Effects on Quality and Customer Satisfaction

	Unresolved Visit Rate		Customer Satisfaction	
	(1)	(2)	(3)	(4)
Treatment	0.002	0.002	0.016	0.016
	(0.002)	(0.002)	(0.014)	(0.014)
Observations	9170	9170	7101	7101
Team FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	Yes

The Unresolved Visit Rate is the proportion of customer service orders that are not resolved during the initial technician visit and require additional follow-up actions or appointments. Customer Satisfaction is measured with a short online survey sent to the customer after a visit. Robust standard errors are shown in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 1.12: Employee and Supervisor Investments in the Development Process

	Number of Trainings	
	(1)	(2)
Treatment	−0.104	−0.102
	(0.140)	(0.139)
Test Participation	0.349***	0.042
	(0.118)	(0.110)
Meeting Conducted	0.229**	−0.121
	(0.111)	(0.140)
Meeting Conducted × Test Participation		0.427**
		(0.190)
Constant	−0.035	0.217
	(0.150)	(0.137)
Observations	797	797

Note. Table shows estimates from linear regressions on the number of trainings per employee. In Column (1) we regress the number of trainings on a treatment dummy, the share of tests an employee has completed during the treatment period, and a binary indicator for employee development meetings that takes value 1 if an employee has indicated the post-experimental survey that they had at least one development meeting in the last 6 months and 0 otherwise. In column (2), we include an interaction term between the meeting indicator variable and the share of tests an employee has completed during the treatment period. Standard errors are clustered at the team level and shown in parentheses. ***, ** indicate significance at the 1% and 5% percent levels, respectively.

Table 1.13: Treatment Effects on Assessment Participation (All Skills)

	Skill 1	Skill 2	Skill 3	Skill 4	Skill 5	Skill 6	All Skills
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	−0.094***	−0.046*	0.001	−0.035	−0.034	−0.147**	−0.066***
	(0.023)	(0.025)	(0.046)	(0.053)	(0.041)	(0.069)	(0.022)
Constant	0.855***	0.793***	0.649***	0.646***	0.897***	0.931***	0.783***
	(0.014)	(0.018)	(0.037)	(0.039)	(0.029)	(0.031)	(0.015)
Observations	2286	2032	1068	450	341	109	2329

Note: Tables shows estimates from linear regressions of each skill assessment during the treatment period on a treatment indicator (columns 1-6) and the average of all post-treatment assessments (column 7). Standard errors are clustered at team level and shown in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 1.14: Treatment Effects on Assessment Participation and Scores

	Assessment Participation		Assessment Scores	
	(1)	(2)	(3)	(4)
Treatment	−0.042**	−0.041**	−0.022	−0.016
	(0.020)	(0.020)	(0.051)	(0.044)
Prior participation	0.376***	0.375***	0.377***	0.341***
	(0.032)	(0.032)	(0.074)	(0.073)
Prior part. × prior score		0.034***		0.597***
		(0.008)		(0.025)
Constant	0.473***	0.475***	−0.354***	−0.313***
	(0.032)	(0.032)	(0.066)	(0.066)
Observations	2232	2232	1988	1988

Note: Table shows estimates from linear regressions on assessment participation (columns 1 and 2) and assessment scores (columns 3 and 4). Participation denotes the number of assignments an employee has taken part in during the treatment period, divided by the number of assignments they have been assigned to. Assessment scores depict the average standardized ratio of correctly answered questions on the knowledge assessment. Prior participation captures the participation rate prior to the experiment. Prior participation × prior score is a multiplicative term of the participation rate prior to the experiment and the average standardized ratio of correctly answered questions in all skill assessments prior to the experiment. Standard errors are clustered at the team level and shown in parentheses. **, *** indicate significance at the 5% and 1% levels, respectively.

Table 1.15: Treatment Effects on Assessment Outcomes by Skill Level

	Participation		Scores	
	(1)	(2)	(3)	(4)
Treatment	−0.046**	−0.093***	−0.022	0.044
	(0.022)	(0.028)	(0.047)	(0.063)
Treatment × High-Skill		0.099***		−0.134*
		(0.027)		(0.073)
High-Skill	0.107***	0.058***	0.772***	0.837***
	(0.014)	(0.017)	(0.037)	(0.052)
Constant	0.738***	0.761***	−0.409***	−0.440***
	(0.017)	(0.018)	(0.037)	(0.042)
p-value High-Skill		0.771		0.109
Observations	2232	2232	1988	1988

Note: Table shows estimates from linear regressions on assessment participation (columns 1 and 2) and assessment scores (columns 3 and 4). Participation denotes the number of assignments an employee has taken part in during the treatment period, divided by the number of assignments they have been assigned to. Assessment scores depict the average standardized ratio of correctly answered questions on the knowledge assessment. The variable 'High-Skill' takes value 1 for all employees with an above median average assessment score prior to the experiment and value 0 otherwise. 'p-value High-Skill' reports the p-value of a Wald test of Treatment + Treatment × High-Skill'. Standard errors are clustered at the team level and shown in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 1.16: Pre-Treatment Characteristics of Survey Participants

	Control (N=503)		Treatment (N=420)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Assessment Participation	0.88	0.24	0.87	0.27	-0.01	0.69
Assessment Score	0.05	0.84	0.01	0.86	-0.04	0.48

Note: Table shows pre-experimental data for employees that participated in the post-experiment survey, split by treatment group. Assessment participation denotes the average participation for the tests every member of the team has been assigned to. Assessment results depict the average standardized ratio of correctly answered questions on a knowledge test. P-values are calculated using OLS, regressing the dependent variable on a treatment indicator.

Table 1.17: Treatment Effects on Job Satisfaction by Employee Skill Level

	Job Satisfaction	
	High-Skill	Low-Skill
Treatment	−0.301***	−0.109
	(0.114)	(0.111)
Constant	0.142*	0.052
	(0.082)	(0.081)
Observations	389	324

Note: Table shows estimates from linear regressions on job satisfaction as reported by the employees in a post-experimental survey. Job satisfaction is measured using a single-item measure from the Linked Personnel Panel (Kampkoetter et al., 2016) and answered on a 7-point Likert scale from 1 = 'Strongly disagree' to 7 = 'Strongly agree'. We split the sample based on the variable 'High-Skill' which takes value 1 for all employees with an above median average assessment score prior to the experiment (column 'High-Skill') and value 0 otherwise (column 'Low-Skill'). Standard errors are clustered at the team level and reported in parentheses. *, *** indicate significance at the 10% and 1% levels, respectively.

Table 1.18: Treatment Effect on Training by Employee Skill Level

	Num. of Trainings	
	(1)	(2)
Treatment	−0.194**	−0.171*
	(0.091)	(0.092)
Treatment × High-Skill		−0.101
		(0.066)
High-Skill		0.115**
		(0.053)
Constant	0.374***	0.348***
	(0.077)	(0.077)
p-value High-Skill		0.062
Observations	2582	2234

Note: Table shows estimates from linear regressions of the number of trainings on a treatment indicator interacted with the 'High-Skill' dummy. The variable 'High-Skill' takes value 1 for all employees with an above median average assessment score prior to the experiment and value 0 otherwise. Standard errors are clustered at the team level and shown in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 1.19: Reasons for Skill Assessment Participation

	Obligation	Demand	Improvement
	(1)	(2)	(3)
Treatment	−0.011	−0.026	−0.101
	(0.063)	(0.069)	(0.081)
Constant	0.005	0.012	0.046
	(0.044)	(0.046)	(0.052)
Observations	850	850	850

Note: Table shows results from linear regressions on questions from the post-experimental survey. Columns capture the perceived obligation to take the test ('I participate in the assessment because it is part of my job.'), supervisor demand ('I participate in the assessment because my supervisor expects me to do it.') and personal improvement ('I participate because it helps me to get better at my job.'). Standard errors are clustered at the team level and reported in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 1.20: Treatment Effects on Meetings (Robustness Checks)

	Prob. of meeting		Num. of meetings	
	(1)	(2)	(3)	(4)
Treatment	−0.106**	−0.131***	−0.148*	−0.160*
	(0.043)	(0.043)	(0.084)	(0.089)
Prior meetings		0.231***		0.444***
		(0.057)		(0.115)
Constant	0.774***	0.539***	1.189***	0.812***
	(0.025)	(0.059)	(0.054)	(0.113)
Level of Analysis	Individual	Team	Individual	Team
Observations	846	95	846	95

Note: Table shows estimates from linear regressions of the likelihood or frequency of development meetings on a treatment indicator. For the individual-level analyses, the dependent variables are a binary indicator that takes value 1 if an employee has indicated to have had more than one meeting with their supervisor (1) and the overall number of meetings (3), which can take values between 0 and 3. For the team level analysis, the dependent variables are the number of employees with at least one meeting per team, relative to the total number of employees per team who responded to the post-experimental survey (2) and the average number of meetings summarized across each team (4). Columns (2) and (4) further include the number of meetings per team as a control variable, which was assessed in the pre-experimental survey. Not every team met the mininum participation threshold of 5 responses, which is why several teams could not be included in this analysis. Re-running the team-level analyses without control variables does not change the results. Robust standard errors (clustered at the team level in columns 1 and 3) in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

Table 1.21: Treatment Effects on Assessment Outcomes and Trainings (Employees with Meetings)

	Participation		Test Scores		Trainings	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	−0.065**	−0.056**	0.024	0.010	−0.121	−0.134
	(0.028)	(0.025)	(0.088)	(0.092)	(0.172)	(0.181)
Prior participation		0.339***		0.423**		0.497***
		(0.066)		(0.195)		(0.164)
Constant	0.844***	0.540***	0.079	−0.283	0.466***	0.057
	(0.019)	(0.064)	(0.050)	(0.180)	(0.111)	(0.132)
Observations	577	565	570	530	614	565

Note: Table shows estimates from linear regressions on assessment participation (columns 1 and 2), assessment scores (columns 3 and 4) and the number of trainings per employee (columns 5 and 6) on a treatment indicator and average pre-experimental participation in the skill assessment. Standard errors are clustered at the team level and shown in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 1.22: Signaling and Reputational Concerns (Controlling for Meeting Incidence)

	Demonstrate Competence		Fear of Incompetence	
	(1)	(2)	(3)	(4)
Treatment	0.053	0.080	0.158	0.152
	(0.208)	(0.204)	(0.184)	(0.184)
Treatment × High-Skill	−0.723**	−0.847***	−0.082	−0.069
	(0.292)	(0.295)	(0.238)	(0.236)
High-Skill	0.262	0.297	−0.658***	−0.662***
	(0.186)	(0.188)	(0.154)	(0.154)
Meeting Conducted	0.956***		−0.094	
	(0.168)		(0.121)	
Number of Meetings		0.496***		−0.072
		(0.073)		(0.051)
Constant	4.038***	4.172***	2.538***	2.553***
	(0.181)	(0.156)	(0.139)	(0.118)
p-value High-Skill		0.000		0.491
Observations	781	781	778	778

Note: Column (1) and (2) show estimates from linear regressions on the strength of an employee's competence motif when considering taking part in the skill assessment. Columns (3) and (4) show estimates from linear regressions on the strength of an employee's avoidance motif when considering not to participate in the skill assessment. All questions are answered on a 7-point Likert scale from 1 = 'Strongly disagree' to 7 = 'Strongly agree'. The variable 'High-Skill' takes value 1 for all employees with an above median average assessment score prior to the experiment and value 0 otherwise. 'p-value High-Skill' reports the p-value of a Wald test of the main effect on high-skill employees. Standard errors are clustered at the team level and shown in parentheses. **, *** indicate significance at the 5% and 1% levels, respectively.

Table 1.23: Treatment Effects on Training by Training Type

	Num. of Trainings		Prop. of Training	
	(1)	(2)	(3)	(4)
Treatment	−0.108**	−0.122**	−0.055**	−0.059**
	(0.051)	(0.055)	(0.024)	(0.026)
Treatment × Skill Preservation	0.015	0.021	0.005	0.006
	(0.065)	(0.070)	(0.027)	(0.029)
Skill preservation	0.024	0.018	0.020	0.018
	(0.054)	(0.059)	(0.021)	(0.024)
Prior participation		0.133***		0.064***
		(0.036)		(0.020)
Constant	0.175***	0.080*	0.088***	0.043*
	(0.044)	(0.045)	(0.020)	(0.024)
p-value: Skill pres.	0.121	0.112	0.091	0.089
Observations	5164	4468	5164	4468

Note: Table shows estimates from linear regressions on the number of assigned trainings per employee. In columns (1) and (2) the dependent variable is denoted by the number of trainings , while columns (3) and (4) cover the probability of training. Skill preservation (or Skill pres.) is a binary indicator that is equal to 1 if the training purpose is the preservation of an existing skill and 0 otherwise. Prior participation captures the participation rate prior to the experiment. Robust standard errors clustered at the team level in parentheses. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

Table 1.24: Treatment Effects by Training Purpose and Employee Type

	Skill Preservation		New Skill Acquisition	
	High-Skill	Low-Skill	High-Skill	Low-Skill
Treatment	−0.095	−0.108*	−0.185**	−0.070
	(0.076)	(0.058)	(0.071)	(0.050)
Constant	0.226***	0.196***	0.235***	0.153***
	(0.057)	(0.051)	(0.066)	(0.040)
Observations	1043	1191	1043	1191

Note: Table shows estimates from linear regressions on training devoted to skill preservation (columns 1 and 2) and new skill acquisition (columns 3 and 4). The variable 'High-Skill' takes value 1 for all employees with an above median average assessment score prior to the experiment and value 0 otherwise. Standard errors are clustered at the team level and shown in parentheses. * p<0.1, ** p<0.05, *** p<0.01

1.8.2 Figures

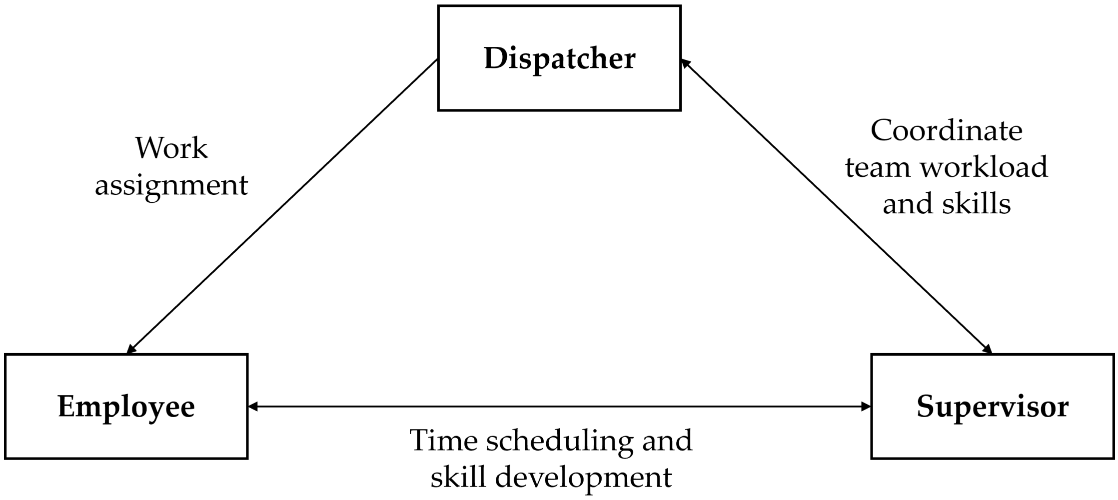

Figure 1.4: Responsibilities in the Firm

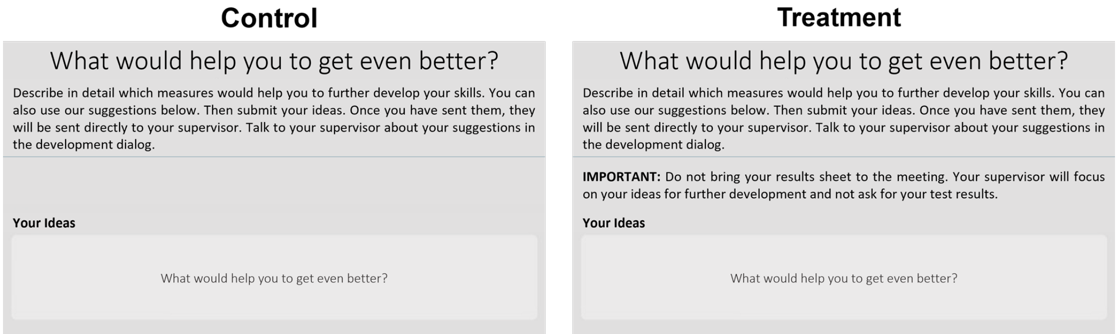

Figure 1.5: Communicating the Treatment on the Skill Assessment Platform

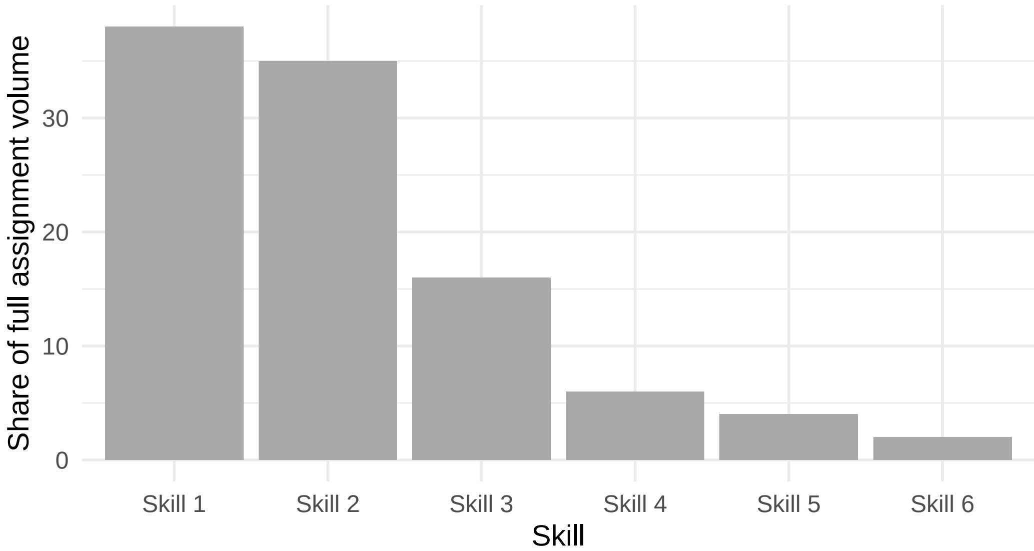

Figure 1.6: Relative Share of Work Assignment Volume per Skill

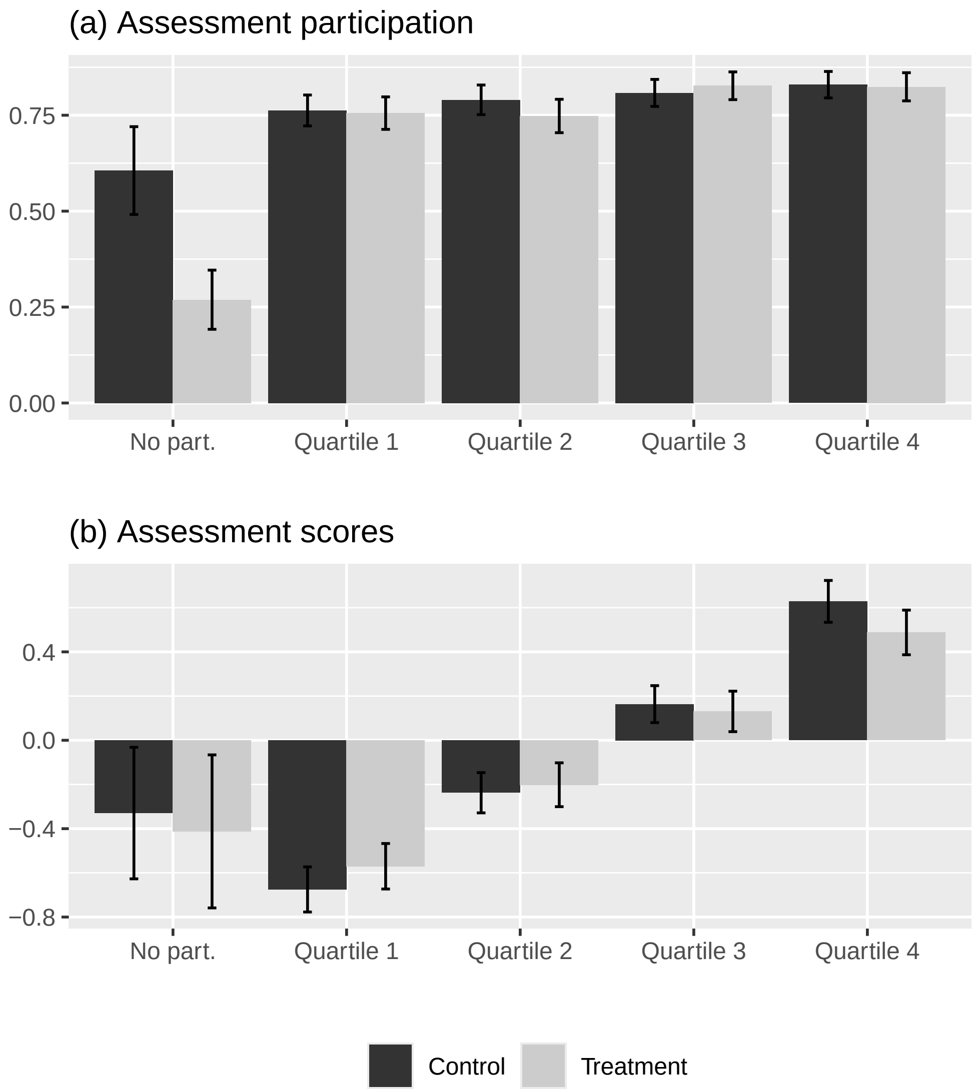

Figure 1.7: Outcomes by Prior Test Participation and Performance

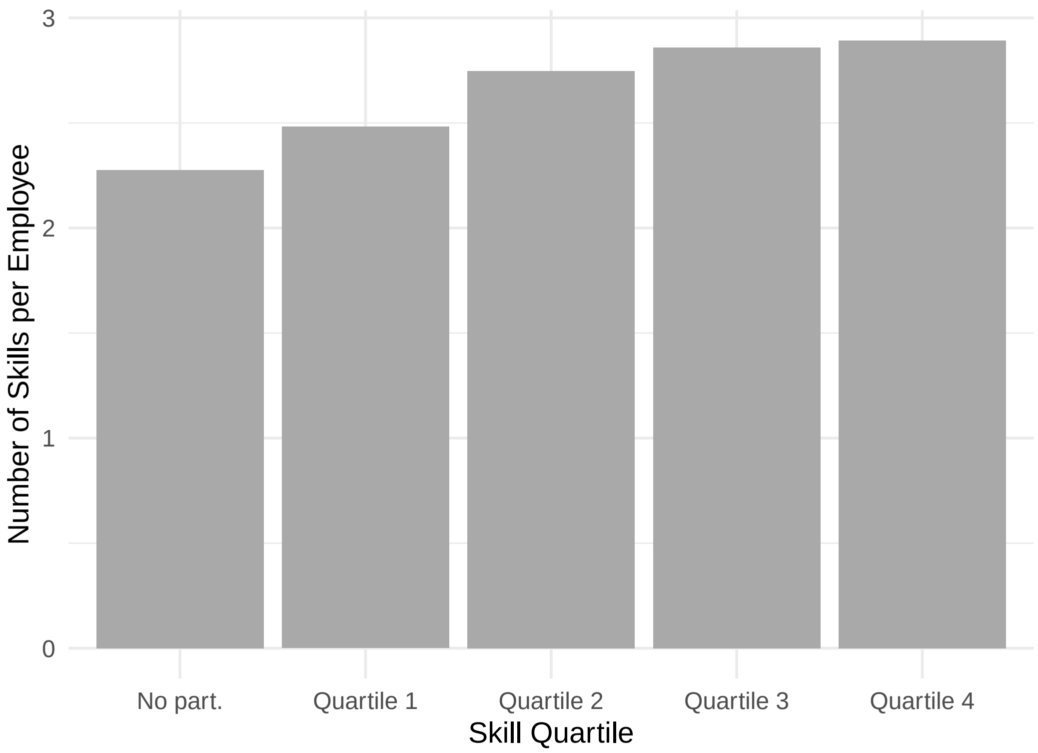

Figure 1.8: Average Number of Skills per Quartile

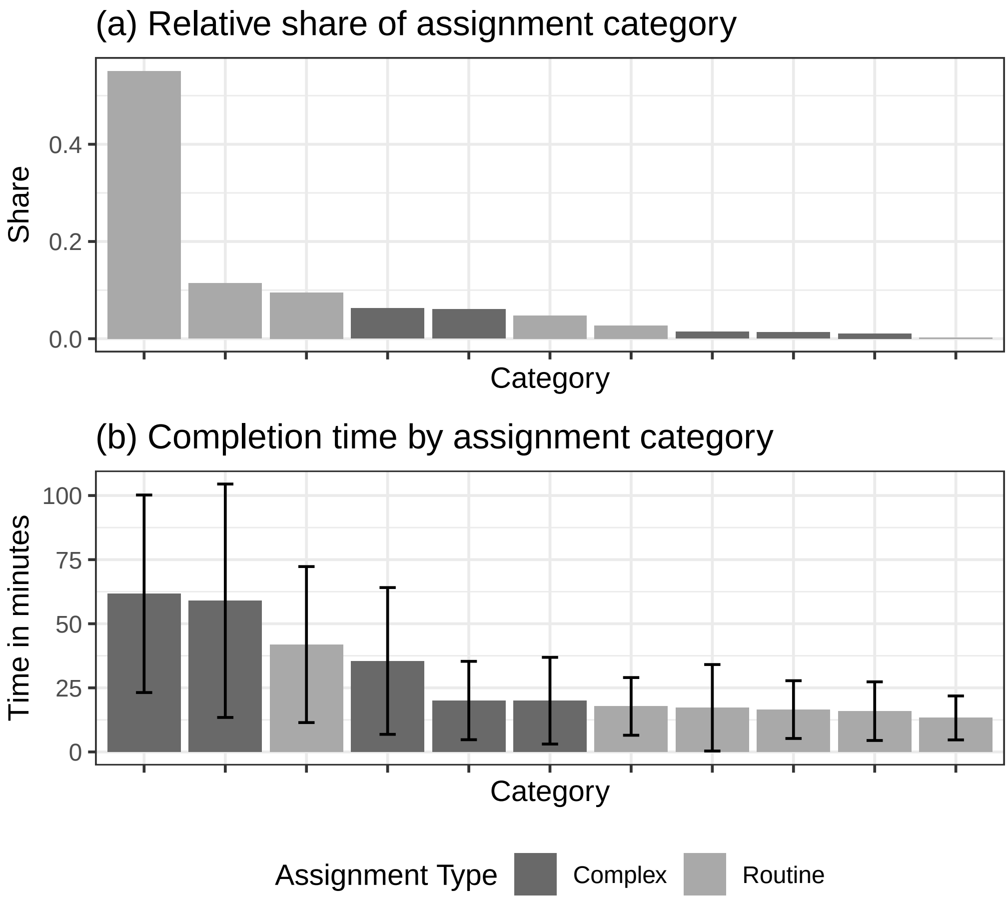

Figure 1.9: Classification of Work Assignments by Complexity

1.8.3 Survey Items (Translated)

Item Text	Name
How useful are the following elements of the learning platform to you? • The skill assessments on the platform • The development ideas you can write down after the skill assessment	*Usefulness*
How easy was the handling of the following elements of the platform? • The skill assessments on the platform • The development ideas you can write down after the skill assessment	*Handling*
I participate in the skill assessment because…	
… it is my job.	*Obligation*
… my manager is expecting it from me.	*Demand*
… it helps me to find out how I can improve.	*Improvement*
… to demonstrate to others that I am competent.	*Competence*
If I do not participate in the skill assessment, it is because…	
… I do not have the time to participate.	*Time*
… the skill is not relevant for my daily work.	*Mismatch*
… I already know that I will achieve a high score.	*Overconfidence*
… I expect to perform poorly.	*Incompetence*
Thinking back to the last 6 months: How often did you have a meeting with your manager to talk about your skill development?	*Number of Meetings*

Item Text	Name
Thinking back to the last 6 months, how do you evaluate the following statements regarding your development process? • My manager helps me find the right training measures. • My manager supports me in learning independently. • My manager is interested in what I am currently learning. • My manager helps me try out new solutions, even if I make mistakes in doing so. • Every time I have a good idea for my own development, my manager implements it. • With regards to my development, it often happened that I was presented with accomplished facts by my manager. • I always had a say in the decisions about my own development.	*Supervisor Support*
How satisfied are you with your work at the moment?	*Satisfaction*
I often think about changing my job.	*Turnover*
• My supervisor gives me special recognition when my work performance is especially good. • My supervisor always gives me positive feedback when I perform well.	*Positive Feedback*
• My supervisor shows me his/her displeasure when my work is below acceptable standards. • My supervisor lets me know about it when I perform poorly.	*Negative Feedback*

<table>
<tr><th>Item Text</th><th>Name</th></tr>
<tr><td>Now we ask you to assess what leadership skills you think your manager has. We have included a short description of a person who has this competence.

• Leadership and coaching: adapts leadership style to the current environment, motivates and mentors others.
• Influence on others: is well connected, uses persuasion and their own authority to achieve team goals.
• Interpersonal relationships: responds appropriately to the needs of others, understands how to give good feedback, builds trust with others quickly.
• Conflict management: anticipates conflicts and complaints and resolves them constructively.</td><td>Leadership Skills</td></tr>
<tr><td>1. I value the professional skills of my manager.
2. I respect my manager’s knowledge and expertise.
3. I am impressed by my manager’s knowledge of the job.</td><td>Perceived Competence</td></tr>
</table>

2 Employee Performance in Response to Workplace Errors: Evidence from the Field

This chapter is based on: Grabe (2025)

Abstract. Errors are an inevitable part of work, and managers play a central role in preventing and addressing them. While a large literature studies the detection and prevention of errors, their consequences for employee performance are not always clear. Using an event-study design with 3,200 technicians across 50 firms, this paper studies how employee performance changes after a manager has confronted an employee with an error identified by an external party. The results show that overall technical performance - as measured by the number of completed work assignments - improves, with no significant impact on assignment quality. However, the intervention by the manager leads to a shift in the employee's priorities: While performance increases in the technical dimension, it decreases in a non-technical dimension. These negative effects are more pronounced after a technical compared to a non-technical error, consistent with errors diverting attention away from non-technical aspects of the job. A back-of-the-envelope calculation suggests that the additional profit from the increase in completed assignments is twice as large as the profit loss associated with lower customer service quality.

Keywords: Errors, Performance, Learning, Multitasking
JEL Codes: J24, M53

This paper is single-authored.

2.1 Introduction

Errors are an inevitable part of every work environment. Many firms therefore have formal processes to hold employees accountable for their actions. These processes are common across business functions –from internal audits in accounting to code review in software development and quality control in manufacturing. These processes are often referred to as *action accountability controls* (Merchant & Van der Stede, 2017) and their primary purpose is the prevention and detection of errors. A large literature has covered the effectiveness of various forms of accountability controls, especially in reducing fraud and misconduct (Altamuro & Beatty, 2010; Ashbaugh-Skaife et al., 2009; Chalmers et al., 2019; Guo et al., 2016). In most cases, line managers are enforcing these controls and it is their job to confront employees with their mistakes and decide on the appropriate consequences. However, as the majority of errors are unintentional and arise from insufficient knowledge or experience (Frese & Keith, 2015; Maletta & Wright, 1996; Rebbitt, 2014), simply informing employees about the desired and undesired actions may not always address the underlying issue. Instead, as echoed by Merchant and van der Stede (2017), managers might take errors as an opportunity to "help provide direction and alleviate personal limitations due to inadequate skills or experience" (p. 92). As a result, exposing and addressing errors may improve performance beyond the prevention of future errors. However, the potential performance effects of action accountability controls are rarely studied, especially from a managerial perspective.

This paper studies how holding employees accountable for their mistakes influences their subsequent performance. Specifically, I focus on how employee performance changes after being confronted with past errors by their managers. For instance, managers might point out areas where employees need to improve and organize or provide training that addresses the skill gaps related to the error (Adhvaryu et al., 2022; Frese & Keith, 2015; Grabe & Sliwka, 2025; Manthei et al., 2023). It is further possible that employees are responding to the increased attention from the management and invest into their skills to avoid negative consequences such as demotion or termination (Brüggen, 2011; De Janvry et al., 2023; Katok & Siemsen, 2011a). However, the increase in attention may also have a downside - as employees are put in an unfavorable light when confronted with a mistake. To make up for the bad impression in the eyes of the supervisor, employees shift their attention towards tasks associated with the error, which may in turn have a negative impact on other areas of the job. Specifically, the intervention by the manager might lead to a shift in the employee's priorities in a way that they neglect other parts of the job (Feltham & Xie, 1994; Hannan et al., 2013; Milgrom & Roberts, 1992). Therefore, while confronting employees with their mistakes should have a positive effect on tasks related to the error, these interventions might have unintended effects in other dimensions unrelated to the error.

To study these effects, I match individual performance data from 3,265 employees in 50 firms with 1,165 detailed error reports, which yields around 167,000 employee-week observations. All firms are service providers for a large technology company (hereinafter referred to as *MultiCo*), whose employees work on the same infrastructure as the service providers. Every day, employees complete around 4-5 work assignments per day where they install and maintain products and services from MultiCo. Employee compensation partially depends on the number of completed assignments and is therefore the main indicator of individual performance. Data on employee errors was obtained from the error management system from MultiCo to which only their own employees can submit a report to whenever they discover an error caused by an employee from a service provider. The report contains details on the nature and potential cause of the error and is submitted to the quality management of MultiCo and the general manager of the service provider. If the report is considered justified, managers have to contact the employee, provide instruction, and document their actions in the error management system.

Using an event-study approach, I examine how employee performance changes after the manager has reached out to the employee regarding an error. I find that employee performance, as measured by the average number of completed work assignments per day, sharply increases around the time point of the manager's intervention. I find that these performance improvements are persistent and primarily driven by employees with lower prior performance. As these employees require more time to reach the plateau of their learning curve, the intervention is more likely to occur when they are still learning on the job. These effects are unlikely to be driven by post-intervention selection effects, differences in the nature of errors, or preferential treatment by the manager. While exposing and addressing skill gaps in response to an error is beneficial for overall performance, I show that these improvements are connected to a reduction in customer satisfaction, but not to the quality of their technical work. These reductions are more likely to occur after technical errors, when employees are more focused on closing skill gaps in the technical rather than non-technical domain of the job.

This paper contributes to the literature on action controls in firms. Action controls are rarely considered from a managerial or behavioral perspective (see Feichter & Grabner, 2020 for a review), even though it is often managers themselves who are holding employees accountable for their actions (Casas-Arce et al., 2024; Heese & Pérez-Cavazos, 2020; Kowaleski et al., 2024). I show that holding employees accountable for their errors contributes to overall employee performance, as additional managerial attention and guidance leads to improve their performance substantially. Similarly, I contribute to a growing literature on the role of managers for employee learning. Prior research shows that managers shape employee development through their use of different management practices such as goal-setting (Arnold et al., 2025), performance feedback (Arnold et al., 2023; Manthei et al., 2023), and uncovering skill gaps (Cai, Chen, et al., 2024; Grabe & Sliwka, 2025). A growing literature illustrates also how managers directly affect employee careers by motivating and mentoring them (Cai, Mahlendorf, et al., 2024; Sandvik et al., 2020, 2025; Wu et al., 2024). I show how managers can contribute to employee development more indirectly by making them aware of their current skill gaps. Finally, I contribute to a broader literature on behavioral responses to organizational policies, such as employee responses to termination notices or wage cuts (Alfitian & Vogelsang, 2026; Cederlöf et al., 2024; Krueger & Friebel, 2022; Sandvik et al., 2021). Contrary to these findings, I find that seemingly "bad news" from an employee perspective can have positive long-term effects on work performance.

The paper proceeds as follows: Section 2.2 lays out the guiding hypotheses of this paper. In Section 2.3, I describe the organizational context and the error management process which constitutes the basis for the analysis, as well as the data and identification strategy for the main analyses. Section 2.4 and Section 2.5 show the corresponding results as well as the exploration of potential mechanisms for the main results. Section 2.7 concludes.

2.2 Theory

This paper focuses on whether and how employee performance changes in response to workplace errors. Following Frese & Keith (2015), I define errors as unintentional – yet avoidable – deviations from specific goals that arise from insufficient knowledge, inadequate judgement, or flawed execution. An error is therefore different than misconduct which requires an explicit malicious intent for personal gain. This definition is similar to existing definitions for related types of errors, such as routine errors (Maletta & Wright, 1996) or mechanical errors (Owhoso et al., 2002; Ramsay, 1994). The behavioral responses to being confronted with an error can be diverse. In the following, I lay out two pathways for how errors can affect performance, one relating to learning from mistakes and the other to avoiding negative consequences.

Errors are often described as "opportunities to learn," and the idea that people can improve their knowledge and skills with the help of errors is based on established literature in behavioral research.[1] For instance, the literature on instrumental learning (Hull, 1943; Metcalfe, 2017; Simon, 1955) even goes as far as stating that "no learning takes place in the absence of errors" (Friston, 2005, p. 1023). The intuition for this claim is that errors provide important information about cause-and-effect relationships. If employees would for instance only memorize work steps, they would lack a deeper understanding of how the individual steps are connected and how they should act if they need to deviate from the original plan.[2] Empirical research further highlights the role of the manager in this process. For example, managers can support learning from errors by triggering reflection and learning or by supporting a culture that facilitates sharing personal insights from errors (Bellora-Bienengräber et al., 2023; Fischer et al., 2018; Frese & Keith, 2015; Van Dyck et al., 2005). As Cronin et al. (2021) show, if managers encourage employees to view mistakes as part of the learning process rather than as a result of their lack of expertise, this increases employee resilience in times of crisis. Similarly, Manthei et al. (2023) present evidence showing that store profit is strongly reduced when store managers no longer discuss problems with their superiors, which again underscores the

[1] For instance, the quote "Failure is just the opportunity to begin again, this time more intelligently" has been attributed to Henry Ford, which emphasizes the value of learning from mistakes. The former COO of Meta Sheryl Sandberg has highlighted the insights generated from failure by saying "We cannot change what we are not aware of, and once we are aware, we cannot help but change".

[2] This further highlights important properties of unintentional errors and intentional misconduct: misconduct does not arise from insufficient knowledge about the task or the set of desired actions, but rather from intentional deviation from these actions. Therefore, misconduct offers no opportunity for learning, as the error does not reveal any deficits in information or ability (see Hamilton & Smith, 2021 for experimental evidence on perceived intention in fraud detection).

role of the supervisor in overcoming challenges at work.

While in all these scenarios, employees learn from their mistakes either on their own or with the support of their managers, this improvement may also be related to the negative attention generated by the mistake. If an error is interpreted by the manager as an indicator that employees have insufficient skills or do not put enough effort into their work, employees might respond by working harder to prove the opposite. In line with this proposition, the literature on impression management and career concerns shows that employees engage in these behaviors for instrumental reasons such as promotions or avoiding termination (Casas-Arce et al., 2023; De Janvry et al., 2023; Katok & Siemsen, 2011a), but also from an intrinsic desire to be appreciated by others, especially their superiors (Bénabou & Tirole, 2006; Casas-Arce et al., 2025; Ewers & Zimmermann, 2015; Grabe & Sliwka, 2025). Based on these potential mechanisms, the following hypothesis is put forth:

H1. Employee performance increases after a managerial intervention in response to an error.

In contrast to the positive effects of managerial confrontation, these interactions can also lead to unintended consequences. One reason for this assumption could be that as managers confront employees with their mistakes, they feel pressured and subsequently focus on the task in which they made the mistake. As prior literature on multitasking and limited attention has shown (e.g. Feltham & Xie, 1994; Hirshleifer & Teoh, 2003; Holmstrom, 1991; Lu, 2022), increasing the salience or perceived value of one task can lead to spillover effects on other aspects of a job. Crucially, for a negative effect to occur, both task dimensions need to be independent from each other and cannot be bases on the same set of skills and abilities. If two tasks share the same inputs meaning that they are connect in the way that allocating more effort to one leads to improvements in the other, the overall effect on both dimensions might even be positive (see Eyring, 2020 for an application in healthcare).

In some cases, these negative spillover effects can be direct consequences of changing management controls, such as performance feedback, incentive design, or monitoring. For instance, Banker et al. (2004) show how the design of performance reports affects employee attention and Hannan et al. (2008) show that receiving precise performance feedback on a specific task increases performance in the target task, at the cost of higher overall performance. In the present setting, this attention shift would not arise from a change in compensation or performance measurement, but rather from the interaction with the supervisor. Employees may - knowingly or unknowingly - adjust their behavior to align it with the perceived expectations of the supervisor (Casas-Arce et al., 2025; De Janvry et al., 2023). This notion is similar to the "effort distortion" effect in Hannan et al. (2013), who show that the desire for self-affirmation and social distinction leads employees to allocate more attention to tasks which help them to stand out from other employees – even at the cost of tasks that contribute more to overall firm performance. Therefore, pointing out errors related to a specific task, might subsequently lead to an over-correction in the employees priorities, which is beneficial for tasks related to the error, but reduces performance in tasks unrelated to the error.

H2. Employee performance increases in areas related to the error, but decreases in unrelated areas.

2.3 Setting and Data

2.3.1 Firm Context

The study sample consists of 50 small-to-medium sized firms in the technical service industry. All companies are subcontractors for a large technology firm, hereinafter referred to as *MultiCo.* Contracts between MultiCo and their subcontractors are negotiated annually and contain target agreements on the volume and desired quality of assignment completion (measured through customer feedback). The key metric behind the subcontractors compensation is the overall volume of completed assignments, however MultiCo can also assign bonuses or penalties depending on whether the firm meets a certain quality target. The average firm size ranges between 10 and 190 employees, with the average firm employing 51 workers who are active for MultiCo customers every year. All employees have to pass a mandatory quality exam before working alongside MultiCo's internal technical service division.[3] It is important to note that even though subcontractors operate in the same technical infrastructure as MultiCo's internal technicians, interaction between employees across or even within firms is very limited, as

[3]Unlike MultiCo employees, subcontractors' employees have not completed a state-recognized apprenticeship as electricians and are often new to the daily work of a service technician. In turn, they receive considerably lower wages compared to employees at MultiCo. Part of an employee's total compensation depends on the number of completed assignments. The exact proportion is at the discretion of the individual companies and is not disclosed to MultiCo or third parties.

most assignments are completed by a single person and team members are distributed evenly across districts and regions. MultiCo's field of work is sub-divided into 37 regions, loosely following German postal codes. Technicians are assigned to one specific region and receive their daily schedule directly from the dispatcher of their own firm. Workers are usually tasked with one of two types of assignments, which are *installations* and *troubleshooting.* In the former case, MultiCo has sold a new service or product to the client and the worker now sets it up for the first time at the client site. In the latter case, a customer reports an issue with an existing product or service. It is then the task of the technician to investigate and provide a solution at the client site and to ensure that all services are working correctly.[4]

2.3.2 Procedures and Data

Table 2.1 reports summary statistics on key variables for all employees who have been active for at least 3 months according to the available data to provide an accurate representation of employee's performance. On average, employees complete 4.5 assignments per day (SD = 1.68) or 19.4 assignments per week (SD = 7.99) and achieve an average star of 4.47 out of 5 stars. Over the entire observation period of three years, employees are on average observed for 87.9 weeks per employee or 1.5 years. Employees receive on average 3.02 error reports over the observation period – however, approximately half of all employees do not receive any error report at all.[5] Taking the number of completed assignments per weekly as a baseline, employees complete around 1040 assignments per year and receive 1.84 error reports, which results in an error rate of 0.17% or 1 in 565 assignments per year. Therefore, while errors are pervasive across employees, the likelihood of receiving multiple error reports is relatively low. Annual turnover across firms ranges between 24% and 32% per year, but the average number of employees per firm remains similar over time, indicating that firms are not facing a net gain or loss in terms of their workforce over time.

In order to support the technicians before and during their visits, both internal and external technicians have access to a logbook that tracks all prior assignments and employees associated with a client or a specific piece of infrastructure. If an employee

[4] The type and complexity of work assignments tends to be distributed homogeneously across firms, as MultiCo's internal technicians cover the larger share of work assignments with a higher degree of complexity (e.g. corporate clients or advanced technologies)

[5] Figure 2.7 shows the distribution of errors across employees. While the likelihood of making at least one error that leads to a report is approximately 50%, the figure shows that the number of errors per employee is not equally distributed throughout this group. As the figure shows, 95% of employees are responsible for only 50% of all errors, while the remaining half can be attributed to just 200 employees.

Table 2.1: Summary Statistics

Variable	M	SD	Q1	Mdn	Q3
Assignments per Day	4.50	1.68	3.45	4.44	5.38
Assignments per Week	19.38	7.99	14.50	19.71	24.25
Star Rating (1-5)	4.47	0.30	4.35	4.51	4.64
Observation Period (Weeks)	87.93	56.23	32.00	83.00	145.00
Total Errors	3.02	5.44	0.00	0.00	4.00

Note: Table reports summary statistics for all employees that have been in the sample for at least 3 months (N = 2405). 'Assignments per Week' and 'Assignments per Day' denote the average number of assignments completed per employee within a given week or day of the observation period. 'Star Rating' reflects the weekly average customer satisfaction score received by the employee from 1 to 5 stars. 'Observation Period' measures the number of weeks between the employee's first and last recorded week with at least one assignment. 'Total Errors' captures the total number of justified error reports an employee received during the observation period.

makes an error during a work assignment, this can be reported through a dedicated reporting system *only MultiCo employees* have access to. The error management process is depicted in more detail in Figure 2.1.

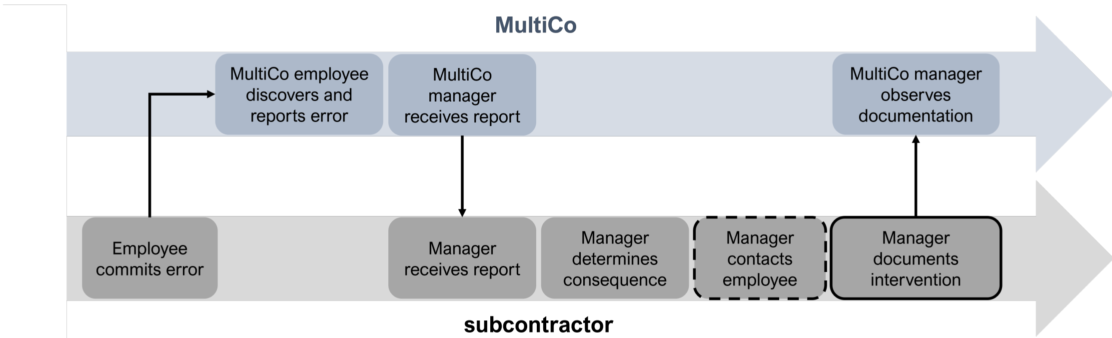

Figure 2.1: Timeline and Structure of the Error Management Process

The errors reported to the manager of the subcontractor usually concern technical issues, such as using wrong materials or impacting the connection of other clients (see Table 2.7). Employees can also report non-technical issues such as insufficient documentation of the visit (e.g. implausible voltage readings). While the initial likelihood that an error occurs will depend on employee characteristics and actions, the discovery of the error depends primarily on the work schedules of internal technicians. Therefore, the specific timing of the error handling by the supervisor is exogenous to the employee who committed the error. The employees discovering and reporting errors do not personally know the employees who commit the error, even though the work in the same infrastructure. It is therefore unlikely that the reporting of the error is influenced by personal relationships. Given that the quality management of MultiCo is also involved in the reporting process, managers are also less likely to treat employees with the same error in a different way.

Upon submission, the report is reviewed by the subcontractor and MultiCo's quality management, who jointly decide whether the error report is justified. On average, 80 percent of errors are considered justified by both parties.[6] Error reports are usually handled directly by the general manager of the subcontractor or a dedicated quality manager. Since the report contains comprehensive documentation and photos, there is limited scope for interpretation at this point. If the error is considered justified by both parties, the manager then contacts the responsible employee and documents the results in the error management system. Managers can choose from different consequences in the documentation, such as on-site training, warnings, and (temporary) suspension. While managers can in theory avoid taking any action, less than 20% of justified errors remain without consequences. Among the available consequences, the most commonly used is remote instruction (81,7%). Onsite training is less frequently used (13.7%), and most firms rarely suspend or terminate employees in response to the first error (4.5%). The latter cases are excluded from the analysis, as they are not errors as defined above. After determining the consequence and taking action, managers are required to submit a statement on how they handled the error. Examples from these statements suggest that managers mostly provide technical instruction (e.g. "*The technician was instructed again on how to conduct the assignment and how to avoid mistakes in the future*" or "*The colleague received verbal training and was reminded again to be more careful when setting up [technology].*").

As Figure 2.1 shows, documenting the details of the intervention is the last step in the error management process. Naturally, employees can only be confronted in the time period between the date of the assessment associated with the error and the date of the documentation by the manager. Based on a sub-sample of 1,029 errors for which the date of the assignment is available, the average processing time is 6.49 weeks (SD = 12.28). As Figure 2.2 shows, most employees can only be confronted close to the time point of the documentation. Specifically, ~40% of all employees are contacted within the last 2.5 weeks prior to documentation, and another 30% is contacted within the last 5 weeks

[6]While MultiCo has an interest in the adequate handling of errors, they are not allowed to influence a manager's decision or contact an employee. However, managers have an incentive to demonstrate to MultiCo that they take these reports seriously and therefore handle them quickly and thoroughly as a sign of proactive compliance.

prior to the documentation. To account for the possibility of an earlier intervention in the empirical analysis, it might be advisable to move the reference period for the event study to an earlier time point, which is discussed in the following chapter.

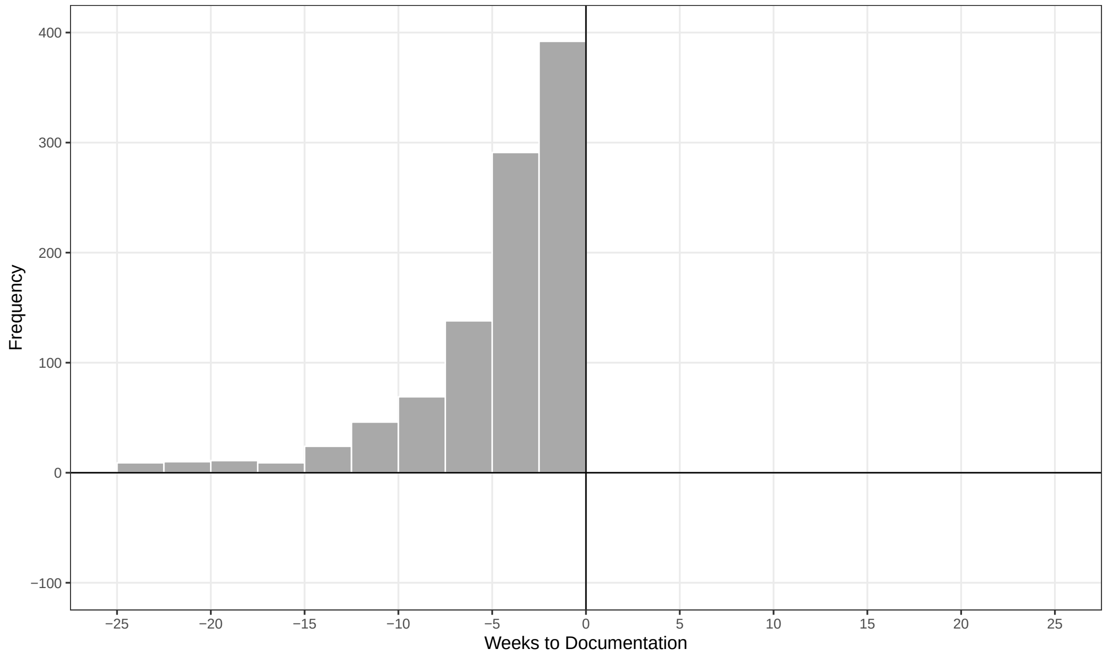

Figure 2.2: Distribution of Time between Error Report and Manager Documentation

2.3.3 Estimation Procedure

To understand how these manager interventions affect employee performance, I use a data set from 3,260 employees in 50 firms, covering the weekly average number of completed assignments over a period of three years, from January 1, 2021, until December 31, 2023, which yields 167,256 employee-week observations. The event study is implemented using the formula Equation 2.1, which represents the standard formula of a two-way fixed effect regression. The key dependent variable is Performance_{it}, which denotes the weekly average number of completed work assignments for MultiCo clients per employee i at time t. I include an individual-specific fixed effect, denoted as Individual_i and a time-specific fixed effect, denoted as Week_t which captures any unobserved effects associated with a given week. The leads and lags relative to the week in which the manager has documented the consequence in the error management system denoted by $\text{Treat}_{it}(k)$. Specifically, $k = 0$ corresponds to the week t when the first manager intervention was documented for employee i. The observed time period ranges from $-26 \geq k \leq 26$ with $k = -26$ capturing all time periods at least 26 weeks before the documentation, and $k = 26$ capturing the average values for 26 weeks after the manager intervention and beyond. Using this approach allows for identification in the absence of never-treated units, as all "not-yet-treated" observations and those outside of the effect window 26 weeks before and after the documentation are used construct the control group (see Schmidheiny & Siegloch, 2023 for a theoretical account on binning in event studies).[7] I use $\text{Treat}_{it}(-7)$ as the reference point of all coefficients, which is the total time frame that 75% of all errors require to be discovered and processed. The variables of interest are the corresponding coefficients for each time period, denoted by β_k. Finally, Error_{it} denotes the idiosyncratic error term.

$$Performance_{it} = Individual_i + Week_t + \sum_{\substack{k=-26 \\ k \neq -7}}^{26} \beta_k \, Treat_{it}(k) + Error_{it} \tag{2.1}$$

2.4 Results

2.4.1 Assignment Performance

Figure 2.3 shows the graphical results of the event study lined out in Equation 2.1. Coefficients and error bars are plotted for each week in relation to the reference period. As the figure shows, no coefficient prior to the reference period is significantly different from zero. After the reference period at $k = -7$, employee performance steadily increases every week – which is in line with a growing share of employees being treated as we move closer to $k = 0$ where all managers have documented their responses to the error (see Figure 2.2). At this point, all employees have been treated and performance has increased by approximately 0.4 assignments (~10%) per day relative to the reference period and remains at a similar level in subsequent periods. To separate the more direct effects of the manager intervention from potential post-intervention effects, I compare the coefficients from the pre-treatment period $k \in [-26, -7]$ to those from the *intervention period* which includes all $k \in [-6, -1]$ and those from the *post-intervention*

[7] Using only not-yet-treated observations as control observations, reduces the risk that the control group consist of employee who are fundamentally different from those who receive error reports. The main analysis therefore does not use the full sample of all 3,265 employees, but the sub-sample of 1,165 employees who have received at least one error report. As Table 2.2 shows, using both the restricted and the full sample does not change the significance or direction of the results.

period $k \in [0, 26]$, respectively. Column (1) in Table 2.2 shows the results for the two periods in relation to the pre-treatment period. In the intervention period, the average number of assignments completed per day increases by 0.254 assignments per day, which is equal to a 5.32% increase relative to the average number of completed assignments prior to the pre-treatment period. For the post-intervention period, the table shows an average performance increase of 0.344 assignments per day, which corresponds to a 8.10% increase.[8] Overall, these results show that employee performance increases substantially during as well as after the intervention by the manager. In turn, this suggests that confronting employees with their mistakes has no negative implications for future performance. Instead, it leads to a persistent and economically meaningful increase in the average number of completed assignments, i.e. the most important performance metric of the firm.

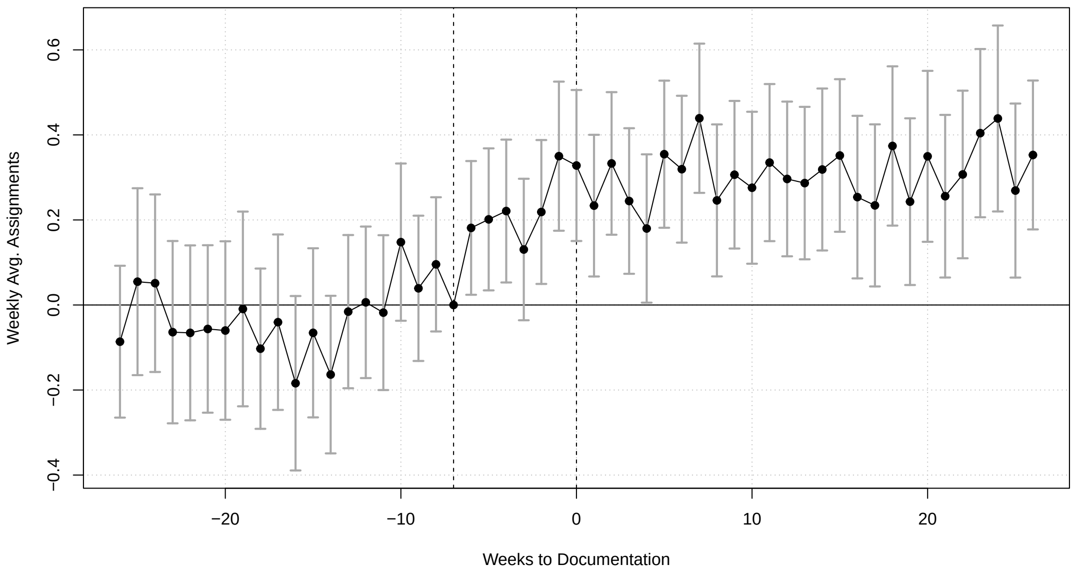

Figure 2.3: Impact of Manager Intervention on Employee Performance

Table 2.2 documents robustness checks for the analyses above. As the table shows, the results are robust to the inclusion of additional region, firm, and year-month fixed effects which control for unobserved heterogeneity in time trends and/or regional characteristics

[8]Column (2) of Table 2.8 shows the combined effect for all time periods after the reference period $k = -7$, while column (3) reports the results for all time periods after the manager's documentation at $k = 0$, which is smaller, but still significant at the 1%-level.

(column 2). Including all untreated employees in the estimation (column 3), and re-running the analysis with a restricted sample of employees who stayed at the firm at least half a year after the intervention (column 4) does not alter the coefficients in a meaningful way.[9] Recent research on the canonical two-way fixed-effect (TWFE) estimator has shown that it can yield biased results, especially when treatment adoption is staggered across multiple time periods (see Roth et al., 2023 for a survey). To address these concerns, I re-run the analysis using an imputation estimator (Borusyak et al., 2024; J. Gardner, 2022; Liu et al., 2024) implemented by Butts and Gardner (2021), which produces similar results (see Figure 2.8 for a comparison between TWFE and 2-Stage DID). Finally, Table 2.10 in the appendix shows that the result is robust to different specifications of the outcome variable, such as different client and assessment types.

2.4.2 Customer Service

While an increase in performance as measured by the number of completed assignments is a positive outcome for the firms, it is also possible that employees achieve these performance gains by compromising on the quality of their work. To evaluate this potential trade-off, the effects on customer satisfaction are examined below. Customer satisfaction ratings are collected in a short survey which is distributed to the customer via text message after the appointment has been completed. Table 2.3 shows the aggregated results for the number of ratings, the number of 5 and 1 star ratings, as well as the average rating. On average, employees receive 3.87 ratings per week with an average rating of 4.49 stars, which is reflected by the high share of 5-star ratings (68.47%) and the rather low share of 1-star ratings (3.61%). As Table 2.3 shows, the overall number of ratings per week as well as the number of 5-star ratings appears to increase slightly, but the potential effect is small and not significantly different from zero. However, as column 3 shows, the number of 1-star ratings increases significantly both during the intervention period (*Treat* × *Intervention*) and afterwards (*Treat* × *Post*). In comparison to the pre-intervention mean, the number of 1-star ratings increases by 0.019 ratings per week in the intervention period (+13.614%), and 0.016 ratings per week in the post-intervention period (+11.188%). Comparing these results to the effect on the overall number of completed assignments (see Table 2.2), the increase in 1-star ratings is almost twice as large. Overall, this negative

[9]As employees might have been confronted by their managers prior to observation period, analyzing the effect on new hires only might provide a more accurate estimate of the true effect of the first confrontation by the manager. Therefore, I re-run the analysis underlying Figure 2.3 for all employees who have been observed in the data set for the first time at least half a year after the beginning of the observation period. As Figure 2.9 shows, the results are similar to those conducted with the full sample.

Table 2.2: Impact of Manager Intervention on Employee Performance

	(1)	(2)	(3)	(4)
Treat × Intervention	0.254***	0.254***	0.196***	0.284***
	(0.052)	(0.052)	(0.051)	(0.062)
Treat × Post	0.344***	0.344***	0.202***	0.384***
	(0.053)	(0.053)	(0.052)	(0.060)
Individual FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	Yes
Region FE	No	Yes	Yes	Yes
Year-Month FE	No	Yes	Yes	Yes
Including Never-Treated	No	No	Yes	No
Excluding Exits	No	No	No	Yes
Firms	34	34	50	32
Employees	1,161	1,161	3,265	841
Observations	95 999	95 999	167 256	84 746

Note: Table shows estimates from two-way fixed effect regressions on the weekly average number of MultiCo assignments per employee per day. Treat × Intervention is a binary interaction term which is equal to 1 for all time periods k between and including -6 and -1, and 0 otherwise. Treat × Post is a binary interaction term which is equal to 1 for all time periods k larger or equal to 0, and 0 otherwise. Column (1) shows results from the baseline specification including only employee- and week-specific fixed effects, where 'week' is a running indicator for the current week in the observation period. Column (2) shows the same estimation including firm-, region, and year-month-specific fixed effects. Column (3) shows results for the estimation in column (2) including all non-treated observations. Column (4) shows results for all employees who are observed in the sample at least 26 weeks after the documentation by the manager. Robust standard errors clustered on the employee-level are shown in parentheses. *, **, and *** indicate significance on the 10%, 5% and 1% level, respectively.

Table 2.3: Impact of Manager Intervention on Customer Satisfaction

	Ratings	5 Stars	1 Star	Mean Rating
Treat × Intervention	0.089	0.042	0.019*	−0.013
	(0.062)	(0.051)	(0.010)	(0.016)
Treat × Post	0.097	0.021	0.016**	−0.022*
	(0.062)	(0.049)	(0.007)	(0.012)
Individual FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Pre Mean	3.874	2.649	0.143	4.495
Firms	32	32	32	32
Employees	971	971	971	971
Observations	59 426	59 426	59 426	58 888

Note: Table shows estimates from two-way fixed effect regressions including individual and week fixed effects.'Ratings' denotes the overall number of customer service ratings a technician receives per week. 'Mean rating' denotes the average number of stars awarded by customers on a scale from 1 to 5, where '5 Stars' and '1 Star' represent the relative share of 5-star, and 1-star ratings across all ratings in a given week. Treat × Intervention is a binary interaction term which is equal to 1 for all time periods k between and including -6 and -1, and 0 otherwise. Treat × Post is a binary interaction term which is equal to 1 for all time periods k larger or equal to 0, and 0 otherwise. Robust standard errors are clustered on the employee-level and included in parentheses. **, * denote significance at the 10% and 5% level, respectively.

effect is reflected in the weekly average rating, which is reduced by 0.022 stars or 0.489% in the post-intervention period.

To put this finding into context, it is important to understand the extent to which the increase in 1-star ratings is related to the quality of the assignment completion. On the one hand, it could be that employees complete more assignments, but many of them require additional rework by MultiCo. In this case, both dimensions are directly tied to each other. On the other hand, if the increase in 1-star ratings reflects the personal interaction with the customer independent of the technician's quality of technical work, the latter dimension should not be affected. As part of the survey, customers are asked not only about their general satisfaction with the technician, but also whether the technician arrived on time for the appointment and whether the installation was successful or the issue was resolved. These evaluation criteria may serve as a more objective measure of the technicians' quality of technical work. Table 2.4 shows TWFE-regressions on punctuality and completion rates, as measured by the customer feedback survey. As the table shows, technicians are on time for 93% of the assignments and complete the assignment to the customers satisfaction in 91% of all cases.[10] While the coefficients shown in columns (1) and (2) of Table 2.4 are negative, the potential effects are small and are not significantly different from zero, which suggests that the quality of work is not affected substantially. To corroborate this finding, a comparison between columns 3 and 4 of Table 2.4 shows that excluding ratings where the assignment was not completed does not change the results. These findings suggest that the reduction in customer service likely reflects poor customer service rather than a reduction in the general quality of the assignments. In other words, confronting employees with their mistakes can be beneficial for performance as measured by the number of completed assignments, but comes at a cost in another, unrelated dimension of their work, which is in line with hypothesis *H2*. To further illuminate the behavioral mechanisms behind these findings, Section 2.5 shows heterogeneity analyses with respect to prior performance, regional characteristics, and error types.

[10]The wording of the questions are as follows: "The assignment was completed within the announced time frame." (punctuality) and "My issue has been resolved successfully." (completion). Both questions are answered with "Yes" or "No".

Table 2.4: Impact on Customer Satisfaction and Assignment Quality

	Assignment Quality		Customer Ratings	
	Technician On Time	Assignment completed	All Assignments	Finished Assignments
Treat × Intervention	−0.003	0.001	−0.014	−0.013
	(0.004)	(0.004)	(0.016)	(0.013)
Treat × Post	−0.004	−0.001	−0.022*	−0.021**
	(0.003)	(0.003)	(0.012)	(0.010)
Individual FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Pre Mean	0.93	0.91	4.49	4.64
Firms	32	32	32	32
Employees	971	971	971	971
Observations	58 636	59 426	58 888	57 844

Note: Table shows estimates from two-way fixed effect regressions including individual and week fixed effects. 'Technician On Time' denotes the share of assignments per week for which the technician arrived at the client's site in the announced time frame. 'Assignment completed' denotes the share of assignments per week for which the technician has completed the assignment in the given time frame according to the customer feedback survey.'All ratings' and 'Ratings Finished' denote the average customer satisfaction rating for all assignments (see column 4 of Table 3) and only assignments which the customer has reported as successfully finished, respectively. Treat × Intervention is a binary interaction term which is equal to 1 for all time periods k between and including -6 and -1, and 0 otherwise. Treat × Post is a binary interaction term which is equal to 1 for all time periods k larger or equal to 0, and 0 otherwise. Robust standard errors are clustered on the employee-level and included in parentheses. **, * denote significance at the 10% and 5% level, respectively.

2.5 Mechanisms

2.5.1 Learning and Reputation

In this section, I explore potential mechanisms for the differential effects on employee performance and the effect on customer service quality. To begin with, I explore the effect of the manager intervention on the number of completed assignments. In particular, I investigate two potential mechanisms lined out in Section 2.2. For one, confronting employees with their past mistakes might lead them to improve their understanding of their work directly or through additional instruction by the supervisor. As a result, employees "learn from their mistakes" and improve their technical skills, which leads to long-term productivity gains. However, it is also possible that the supervisor views the error as a signal of poor performance. In this case, employees may respond by working harder to restore their reputation or to avoid demotion or even termination.

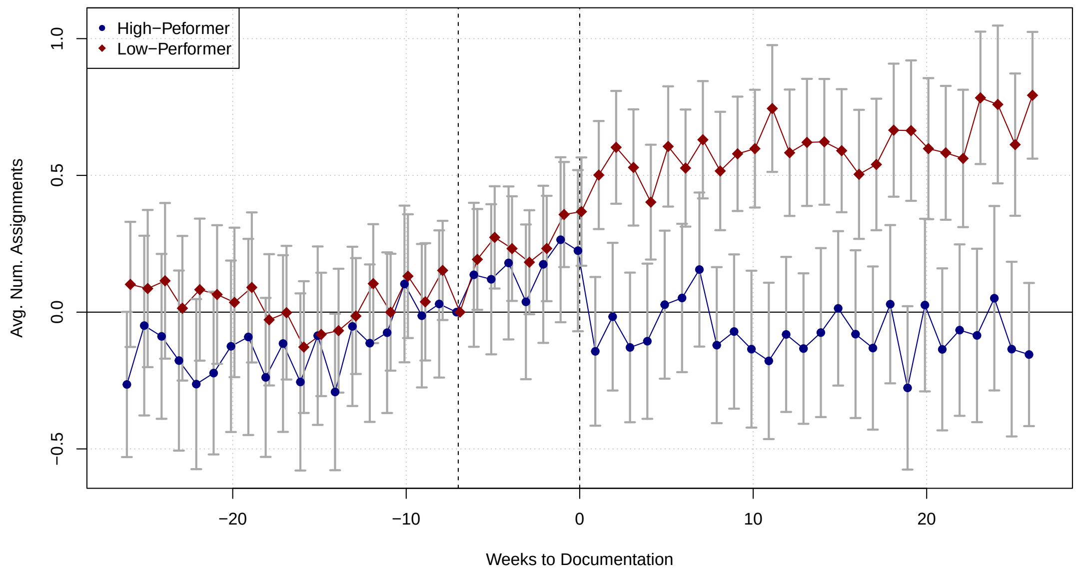

Figure 2.4: Effect of Manager Intervention by Prior Performance

As Figure 2.4 shows, while the performance of both high- and low-performers increase in the intervention period, the post-intervention performance differs substantially between both groups. On the one hand, low-performing employees improve substantially, reaching their highest performance a few weeks after all employees have been treated.

On the other hand, the number of completed assignments among high-performing employees is sharply reduced once every employees has been treated. In line with this observation, Table 2.5 shows that performance in both groups increases significantly during the intervention period, but only the effect among low-performing employees persists during post-intervention period. Therefore, it is conceivable that low-performers improve their technical skills in response to the error and become more productive in the longer term. If anything, the results in Figure 2.4 and Table 2.5 suggest high-performers rather than low-performers have the desire to restore their reputation and therefore work harder when they are under scrutiny by the manager. This interpretation would be in line with the results from Grabe and Sliwka (2025) who show that highly-skilled employees have a stronger desire to demonstrate their skills to their supervisor. It is further possible that long-term improvements among high-performing employees are less likely to happen, because it is generally harder for them to improve even further. However, as column 2 in Table 2.5 suggests, high-performers might become more efficient in response to the confrontation by the manager, as they require marginally less time to complete the same number of assignments as in the pre-intervention period. Therefore, it might be possible that not only low-performers learn from the additional interaction with the manager. Overall, the manager intervention leads to an improvement in employee performance, which is in line with *H1* – and a reason for this could be that these interventions serve as an impetus for low-performing employees to fill existing gaps in their current skill set.

Table 2.5: Impact of Manager Intervention by Prior Performance

	High-Performers		Low-Performers	
	Assignments	Days Worked	Assignments	Days Worked
Treat × Intervention	0.345***	0.044*	0.151**	0.037
	(0.082)	(0.026)	(0.063)	(0.027)
Treat × Post	0.096	−0.044*	0.525***	−0.016
	(0.075)	(0.024)	(0.073)	(0.026)
Individual FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Pre Mean	5.83	4.33	4.48	4.23
Firms	29	29	34	34
Employees	469	469	632	632
Observations	46 434	46 631	45 117	45 303

Note: Table shows estimates from two-way fixed effect regressions including individual and week fixed effects on the weekly average number of assignments per day (column 1 and 3), and the number of days per week with at least one completed assignment (columns 2 and 4), respectively. 'High-Performer' includes all employees with above median average performance prior to the error, as measured by the number of completed assignments. 'Low-Performer' includes all employees with below-median scores according to this metric. Treat × Intervention is a binary interaction term which is equal to 1 for all time periods k between and including -6 and -1, and 0 otherwise. Treat × Post is a binary interaction term which is equal to 1 for all time periods k larger or equal to 0, and 0 otherwise. Robust standard errors are clustered on the employee-level and included in parentheses. *, **, and *** indicate significance on the 10%, 5% and 1% level, respectively.

To further corroborate this interpretation, it is useful to investigate relationship between the learning curves of high- and low-performing employees and the timing of the error alongside this curve. Figure 2.5 shows the average number of completed assignments per day over the first full year of employment. As the figure shows, high-performing employees have a substantially steeper learning curve than low-performers. That is, they are more likely to reach the plateau of their learning curve already after a few weeks. Given that low-performers require more time to finish learning on the job, they are also more likely to receive their first error report while they are still learning on the job. For this reason, the intervention might be particular effective for them, as it provides them with an opportunity to focus on areas where they need to improve. In line with this interpretation, Table 2.9 in the appendix shows that low-performing employees receive their first error report slightly earlier than their high-performing counterparts (39.41 weeks vs. 44.31 weeks, $p = 0.050$). However, the table does not suggest that high- and low-performers differ in the types of errors they make or in the response they receive from the manager.[11]

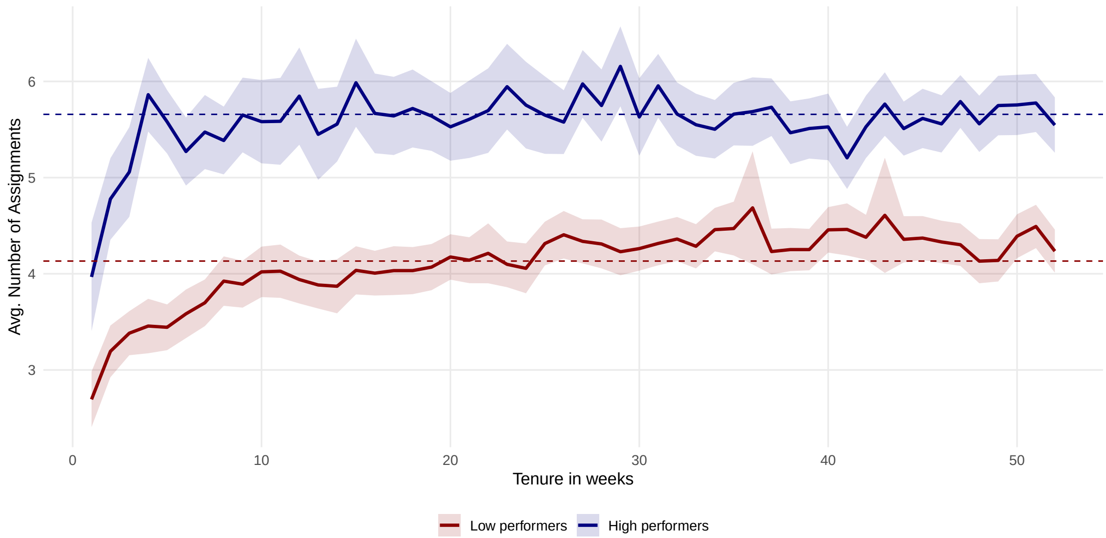

Figure 2.5: Learning Curves by Prior Performance

While low-performing employees might generally be more inclined to learn from their mistakes, they are also at a higher risk of being demoted or terminated. Therefore, they may work harder or have a stronger incentive to learn. As the literature in labor economics shows, in regions with higher unemployment it becomes more attractive to replace workers rather than providing training to them (Arellano-Bover, 2022; Brunello, 2009; Majumdar, 2007). Therefore, employees in regions with higher unemployment may be at a greater threat of being replaced due to bad performance. To explore heterogeneity in the role of this risk, I combine German census data on local unemployment (DeStatis, 2025) with the regions provided by MultiCo. While employees can take on assignments in different regions, they are usually assigned to a single region which is why they are

[11]To study differential attrition, I compare the average turnover rate 12 weeks after the manager intervention with all time periods before that. As Figure 2.10 shows, turnover does not seem to increase after the intervention relative to later time points. If anything, turnover is decreasing among low-performers in the first few weeks after the intervention.

allocated to the regions they completed the most assignments in. I then categorize each regions as “high” or “low” unemployment using a median split. Figure 2.6 shows the results from the main analysis for all employees in each group. As the figure shows, there are no discernible differences between employees who work in regions with above (red line) or below median (blue line) unemployment during the observation period. It is therefore unlikely that the fear of termination is the primary force behind the performance increases among low-performing employees.[12]

Figure 2.6: Effect of Manager Intervention by Regional Share of Unemployment

[12]Restricting the sample to low-performing employees suggests that - if anything - employees who work in regions with *lower* unemployment are more likely to respond to the intervention (see Figure 2.11).

2.5.2 Multitasking

To study the attention trade-off between assignment completion and customer service, I investigate heterogeneity in the effect with respect to the type of errors employees make. To do so, I split the sample by non-technical and technical errors. Technical errors are the result of insufficient technical knowledge or flawed execution, such as using the wrong technique or materials or accidentally impacting another customers service. Non-technical errors on the other hand cannot be traced back to a clear deficiency in the technical domain and therefore include cases such as providing the wrong information to the customer or insufficient documentation after a customer visit. Among the 1,165 first errors covered in the analysis, around 58% are technical errors and the remaining 42% can be considered non-technical errors (see Table 2.7 in the appendix for further details). While technical errors might offer a better opportunity for learning, because the source of error is more precisely defined, they might also attenuate the focusing effect whereby employees shift attention away from non-technical aspects of the job.[13] To test this conjecture, Table 2.6 shows results for technical and non-technical errors with respect to the average rating and the number of 1-star ratings. To begin with, the pre-intervention means for employees in both groups are remarkably similar, which suggests that this analysis in not merely capturing underlying differences between the two groups.[14] As the table shows, the negative effects on customer satisfaction are indeed concentrated on technical errors, as the number of 1-star ratings increases significantly for both the intervention as well as the post-intervention period. This effect is absent for non-technical errors, suggesting that in these cases, employees are not as likely to adjust their priorities on the job.

[13] If technical errors provide a better opportunity for learning, the effect on employee performance as measured by the number of assignments should be larger for this group as well. In support of this finding Figure 2.12 suggests the effect for technical errors is larger on average compared to non-technical errors.

[14] To further support this observation, Table 2.11 reports results from a regression of different employee characteristics (e.g. prior performance, prior tenure, overall tenure, total number of errors) on a binary indicator which is equal to 1 if the first error made by the employee can be categorized as technical, and 0 otherwise. The results show no significant relationship between the indicator and the employee characteristics. The corresponding p-values of the F-Test are 0.129 excluding and 0.307 including firm- and region-specific fixed effects.

Table 2.6: Impact of Manager Intervention on Customer Satisfaction by Error Type

	Technical Errors		Non-Technical Errors	
	Mean Rating	1 Star	Mean Rating	1 Star
Treat × Intervention	−0.024	0.032**	0.006	−0.004
	(0.019)	(0.013)	(0.028)	(0.016)
Treat × Post	−0.027*	0.018**	−0.016	0.013
	(0.014)	(0.009)	(0.021)	(0.011)
Individual FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Pre Mean	4.49	0.14	4.5	0.15
Firms	31	31	28	28
Employees	657	657	314	314
Observations	38 751	39 101	20 137	20 325

Note: Table reports estimates from two-way fixed effects regressions including individual and week fixed effects. The dependent variables capture different dimensions of customer satisfaction: Mean rating denotes the average feedback score as measured by the star rating from 1 to 5 stars (column 1 and 3), while 1 Star denotes the number of 1 star ratings in a given week (columns 2 and 4). 'Technical Errors' encompasses all employees with a technical error according to Table A1 while 'Non-Technical Errors' encompasses all employees who commtted a non-technical error according to Table A1. Treat × Intervention captures the average effect during the intervention period (weeks –6 to –1), and Treat × Post captures all weeks thereafter. Standard errors are clustered at the employee level and reported in parentheses. **, * denote significance at the 10% and 5% level, respectively.

2.6 Cost-Benefit Analysis

Finally, it is important to consider whether the positive effects on the number of assignments outweigh the costs of the reduction in customer service quality. To calculate the magnitude of these countervailing effects, I use proprietary data on assignment revenue and profits per MultiCo customer, and combine these data with industry benchmarks on cost structures in the technical service industry. To begin with, the profit increase for the subcontractor is derived from the post-intervention increase in assignments shown in Table 2.2 multiplied by the estimated profit per assignment which is derived from the revenue per assignment and the firms' costs structure.[15] The compensation per assignment varies across firms and cannot be disclosed publicly. Instead, all monetary amounts are expressed in relation to the estimated profit from one standard assignment. While the additional profits associated with the increase in assignments completed are realized directly by the subcontractor, the decrease in customer service quality is more likely to affect MultiCo, as customers may be reluctant to extend their contract or cancel it entirely if possible. The estimated profit loss can be calculated from the post-intervention increase in 1-star reviews in Table 2.3, the probability of a customer cancelling their contract in case of a 1-star rating and the total profit per customer for MultiCo over the entire customer life-cycle (which was obtained from internal data from MultiCo). Even though MultiCo cannot pass on these costs directly to the subcontractor, they might account for such risk in future negotiations, for example by reducing the amount of assignments or offering less revenue per assignment. Assuming that MultiCo passes on these costs to the subcontractor, the overall net profit from the intervention can therefore be estimated as follows:

$$\Delta\ \text{Profit} = \underbrace{\text{Effect}_{\text{assignment}} \times \text{Profit}_{\text{assignment}}}_{\text{Profit Gain from Effect on Assignments}} - \underbrace{\text{Effect}_{\text{rating}} \times P(\text{cancel} \mid \text{1-star}) \times \text{Profit}_{\text{client}}}_{\text{Profit Loss from Effect on 1-Star Ratings}} \tag{2.2}$$

To illustrate the calculation behind Equation 2.2, the profit per assignment is set to 10 USD, which would lead to an additional profit of 13.73 USD per week per employee.

[15]Since the majority of companies in the sample are not required by law to publish their income statements, cost structures cannot be estimated individually but must be approximated based on local industry standards. I therefore approximate the average cost ratio per unit of output using administrative data on the local construction and installation sector. The margin factor, which determines the profit per assignment, is therefore defined as the the net operating surplus minus personnel costs, divided by the total revenue per assignment.

While a 1-star rating reflects a strong dissatisfaction with the service, the switching costs in the industry MultiCo operates in are relatively high. As a result, a two-fold increase in the likelihood to cancel the contract would yield an estimated profit loss of around 400 USD. Given that the increase in 1-star reviews in the post-intervention period is equal to 0.016 ratings per week, the associated profit loss for MultiCo would be equal to 7.301 USD per week per employee. In summary, even though the reduction in customer service quality is economically meaningful, the profit associated with the additional assignments is around twice as large.

2.7 Conclusion

I investigated the impact of error management on employee performance in a field study of 50 small-to-medium-sized firms in the technical service industry. I find that managerial attention in response to an error leads to a significant increase in employee performance. The effects are largest for employees with low prior performance who are more likely to be contacted by the manager while they are still learning. In contrast to that, further analyses show that the performance increase comes at the cost of lower assignment quality. Comparing intervention effects for technical and non-technical errors suggests that technical errors divert attention away from the customer service dimension of the job. Overall, these findings illustrate that critically engaging with errors can have positive effects and highlight the central role of managers in this process. This paper has important limitations: For one, by nature of the event-study approach, I cannot credibly exclude all alternative explanations for the demonstrated effects. Even though the results are robust to different samples and estimation windows, it is possible that the manager intervention might be a result of their learning process rather than the cause, as low-performers have both lower tenure and receive their first error report at an earlier time point. Additionally, I do not observe more detailed characteristics of employee performance (e.g. the average completion time) or employee perception of the manager intervention. Future research should investigate organizational policies in response to errors in different settings. For instance, it would be interesting to explore the impact of personnel consequences in settings where errors are more common, such as in creative or problem-solving tasks. Finally, it is important to consider employee perceptions more directly. While the results of this study suggest that personnel consequences can have a positive impact on employee performance, I cannot verify whether employees perceive these consequences as positive or negative. Additionally, it would be valuable to learn more about between-firm differences in error management and whether some approaches are considered superior to others. With regards to the managers themselves, it would be interesting to understand their rationales when confronting employees with their mistakes. Even though I do not find systematic differences between high- and low-performing employees, it is likely that managers approach these decisions in different ways. Understanding these factors could provide valuable insights into the effectiveness of error management and action accountability controls in general.

2.8 Appendix

2.8.1 Tables

Table 2.7: Relative Share of Error Types

Error Description	Total	Relative Share
Technical Errors	**676**	**58.06%**
Wrong technique or materials used		21.41%
Incorrect connection established		18.83%
Implausible measurements		11.84%
Impacted connection of other customers		5.96%
Non-Technical Errors	**489**	**41.94%**
Unspecified source of error		23.55%
Insufficient or wrong assignment documentation		7.91%
Incorrect information provided to customer		5.70%
Insufficient customer service		3.15%
Assignment not processed correctly		1.63%
Total	**1,165**	**100.00%**

Table 2.8: Impact of Manager Intervention on Employee Performance

	Seperate Analysis	Including Intervention	Excluding Intervention
Treat × Intervention	0.254***		
	(0.052)		
Treat × Post	0.344***	0.321***	0.263***
	(0.053)	(0.048)	(0.045)
Individual FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Firms	34	34	34
Employees	1,161	1,161	1,161
Observations	95 999	95 999	95 999

Note: Table shows estimates from two-way fixed effect regressions including individual and week fixed effects. In column 1, the interaction term Treat × Intervention depicts the average of all coefficients for the time-periods from -6 until -1, while Treat × Post encompasses all time periods from 0 onwards. In column 2, Treat × Post is a binary indicator that is equal to 1 for all time periods k larger than -7, and 0 otherwise. In column 3, Treat × Post is a binary indicator that is equal to 1 for all time periods k larger or equal to 0, and 0 otherwise. *** indicates significance at the 1% level.

Table 2.9: Differences between High- and Low-Performing Employees

	High-Type (N=464)		Low-Type (N=611)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
Observation Period	123.46	48.33	99.54	52.62	-23.91	<0.001
Weeks to Error	44.36	35.81	39.42	36.38	-4.94	0.046
Share Technical Errors	0.68	0.47	0.70	0.46	0.02	0.406
Share Intervention	0.82	0.39	0.82	0.38	0.00	0.850
Share Remote Instruction	0.64	0.48	0.63	0.48	-0.01	0.669
Share Onsite Instruction	0.06	0.23	0.08	0.27	0.03	0.094
Share Warning	0.12	0.33	0.11	0.32	-0.01	0.679
Statement Length	160.69	196.30	127.40	165.32	-33.29	0.008
Share Standard Statement	0.42	0.49	0.51	0.50	0.10	0.005

Note: Table shows summary statistics for different characteristics between employees with above-median (High-Type) and below-median (Low-Type) average performance. Observation period represents the number of weeks an employee is observed in the data set, while Weeks to Error denotes the number of weeks an employees has been observed in the data set before receiving their first manager intervention. For the types of errors, Share Technical Errors denotes the relative share of errors related to technical errors compared to non-technical ones. Share Intervention shows the share of error reports with a documented response, while Remote Instruction, Onsite Instruction, and Warning show the relative share for the two most common types of responses documented by the manager. Finally, statement length represents the character length of the statement in the error management system describing the actions of the manager and Standard description denotes a dummy variable that is 1 if the manager selected the pre-filled statement of the error management system and 0 otherwise.

Table 2.10: Impact of Manager Intervention on Alternative Measures of Performance

	Client Type			Efficiency	
	MultiCo	Non-MultiCo	All Clients	Sum Assignments	Days Worked
Treat × Intervention	0.254***	0.259***	0.476***	1.305***	0.039**
	(0.052)	(0.061)	(0.083)	(0.255)	(0.019)
Treat × Post	0.344***	0.124*	0.442***	1.373***	−0.033*
	(0.053)	(0.067)	(0.089)	(0.263)	(0.018)
Individual FE	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes
Pre Mean	4.75	4.75	9.49	41.06	4.28
Firms	35	35	35	35	35
Observations	95 999	95 999	95 999	95 999	96 383
Employees	1,161	1,161	1,161	1,161	1,161

Note: Table shows estimates from two-way fixed effect regressions on different types of assignments per employee per day. In columns 1 and 2 results are differentiated with respect to MultiCo and Non-Multico clients, while column 3 shows results for all assignments regardless of client. Column 4 shows results for the weekly sum of all assignments. Days worked represents the share of days with a non-zero number of assignments in a work week from Monday to Friday relative to the days with 0 completed assignments. Treat × Intervention is a binary interaction term which is equal to 1 for all time periods k between and including -6 and -1, and 0 otherwise. Treat × Post is a binary interaction term which is equal to 1 for all time periods k larger or equal to 0, and 0 otherwise. Robust standard errors are clustered on the employee-level and included in parentheses. *, **, ***, indicate significance on the 10%, 5% and 1% level.

Table 2.11: Predicting Technical Errors from Employee Characteristics

	(1)	(2)
Prior Performance	−0.013	−0.015
	(0.010)	(0.012)
Tenure	0.000	0.000
	(0.000)	(0.000)
Weeks to Error	0.000	0.000
	(0.001)	(0.001)
Total Errors	−0.005	−0.006
	(0.003)	(0.004)
Firm FE	No	Yes
Region FE	No	Yes
Observations	863	863

Note: Table shows results from linear regressions on a binary indicator which is equal to 1 if the first error can be caterogized as technical and 0 otherwise. Prior performance denotes the average number of completed assignments per week prior to the intervention of the manager. Tenure denotes the total number of weeks an employee is observed in the observation period. Weeks to error represents the number of weeks in the observation prior to the intervention. Total errors includes all errors recorder for an individual employee.

2.8.2 Figures

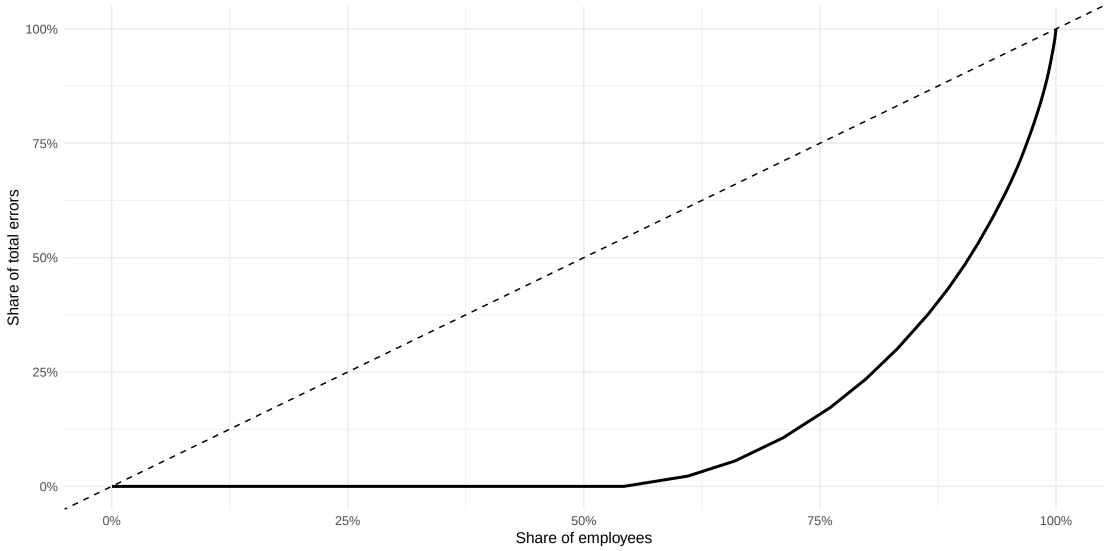

Figure 2.7: Concentration of Errors Across Employees

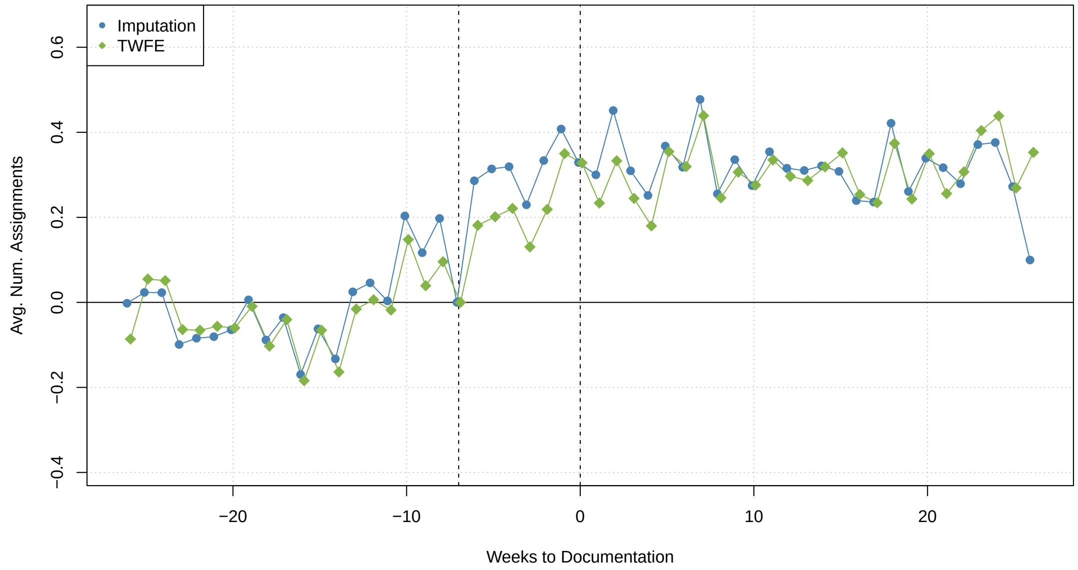

Figure 2.8: Impact of Manager Intervention on Employee Performance (Imputation Estimator)

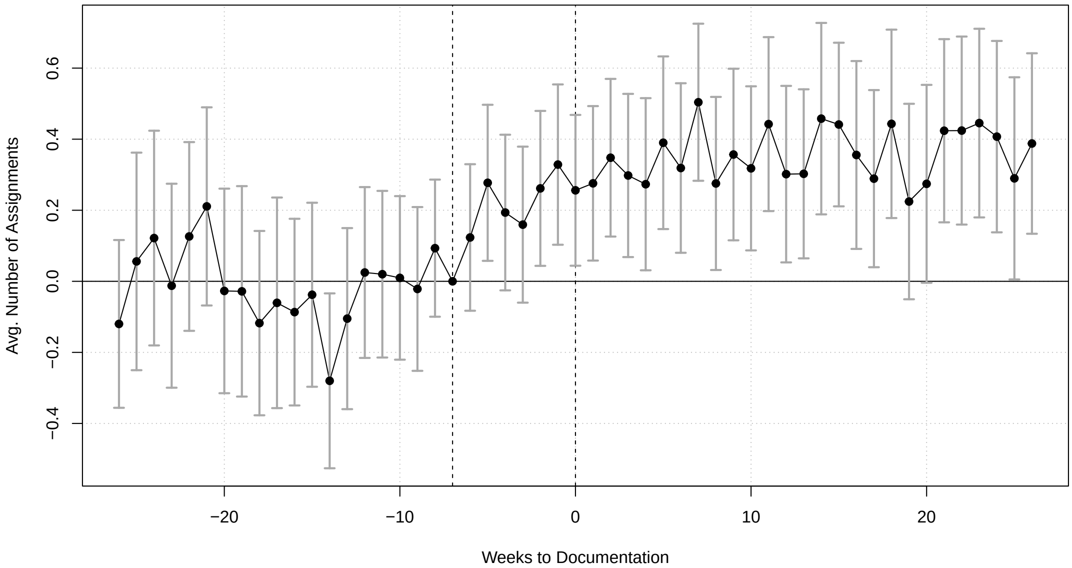

Figure 2.9: Impact of Manager Intervention on Employee Performance (Only New Hires)

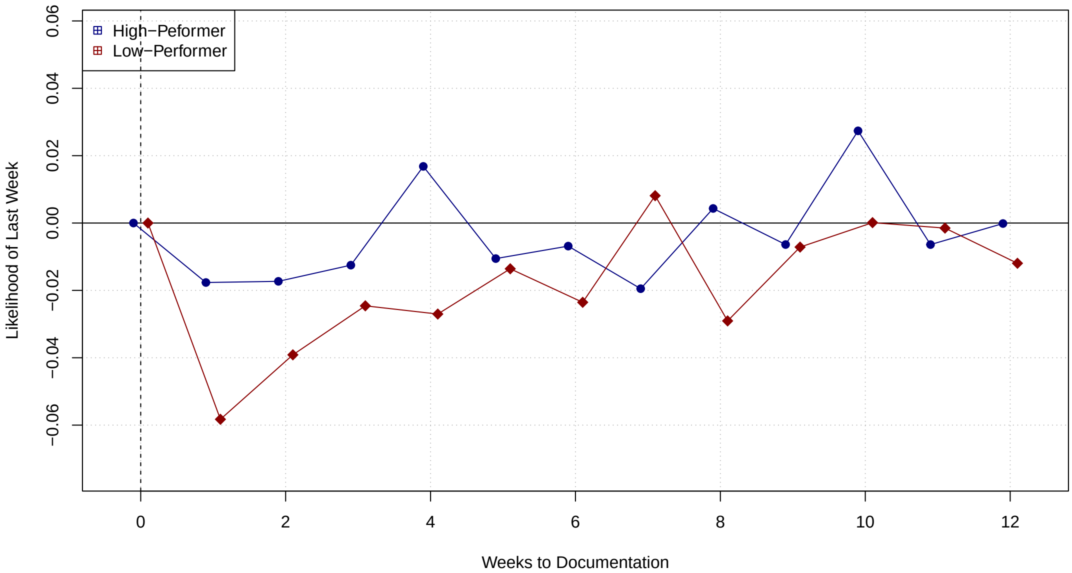

Figure 2.10: Post-Intervention Turnover relative to Group Baseline

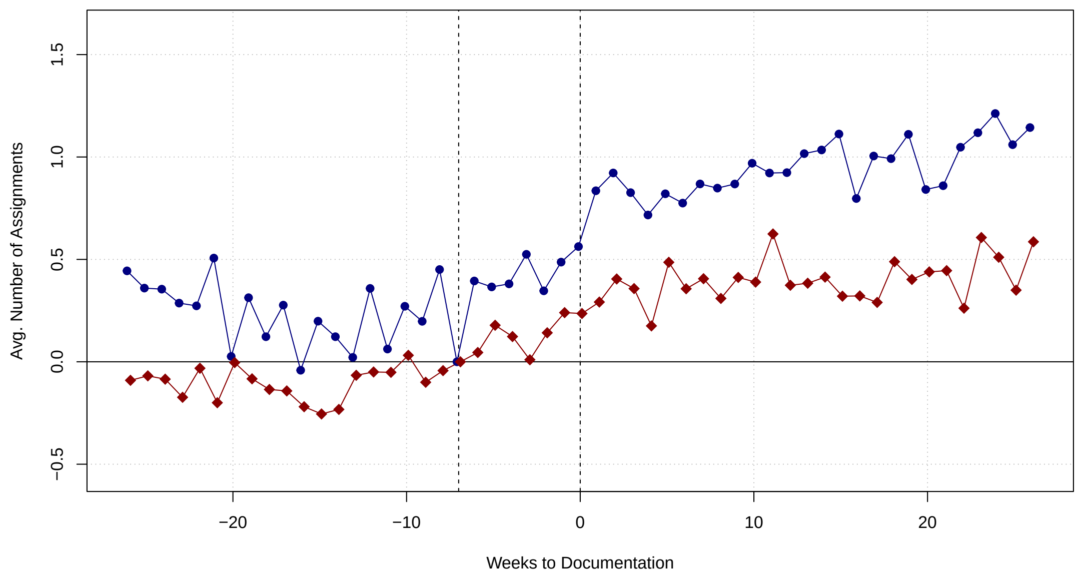

Figure 2.11: Manager Intervention and Regional Unemployment (Only Low-Performers)

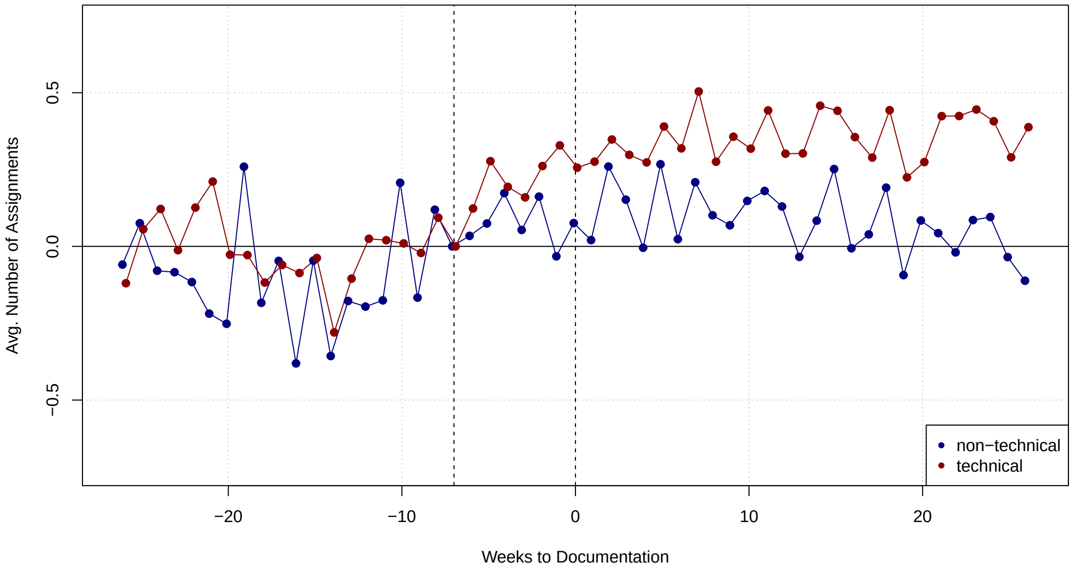

Figure 2.12: Response to Intervention by Error Type

3 Shaping Habits in Organizations: A Field Experiment

Abstract. We investigate whether organizations can shape work habits through monetary incentives. In a field experiment with 829 service technicians in 15 firms, we randomly allocated half of the technicians in each firm to a treatment group receiving bonuses for regularly performing sales activities for 12 weeks. We find a significant increase in sales activities not only during but also after the incentive phase. Using data from a post-experimental survey, we compare different behavioral channels. We find no evidence for increased automaticity, human capital acquisition, or signals about task priorities, but strong evidence for the role of acquired taste: Technicians in the treatment group report higher levels of intrinsic motivation to perform sales activities even after the temporary incentive period has ended.

Keywords: Habit Formation, Taste Acquisition, Temporary Incentives
JEL Codes: M12, M52, C93

This paper is co-authored with Jun.-Prof. Dr. Saskia Opitz from the WHU Otto-Beisheim-School of Management and Prof. Dr. Dirk Sliwka from the University of Cologne. The concept of the experiment was jointly developed by all authors. I was responsible for the implementation of the experiment, as well as the processing and provision of confidential data. I designed the post-experimental survey with support from Jun.-Prof. Opitz, which we then jointly implemented. Jun.-Prof. Opitz was responsible for the analysis of the experimental and survey data. The first version of the draft was developed by Jun.-Prof. Opitz and me, which was then jointly revised by all authors.

3.1 Introduction

Empirical evidence shows that job tasks evolve substantially over time as firms adopt new technologies, adjust workflows, and respond to competitive pressures, requiring employees to accommodate new tasks and take on additional responsibilities (Autor et al., 2024; Gibbs & Bazylik, 2022). Even when firms succeed in initiating short-run adjustments in employee behavior, sustaining these changes over time often remains a major challenge. A key reason for this is that lasting change requires behaviors to be integrated into employees' routines. Psychological research indeed suggests that nearly half of an individual's daily actions can be considered 'habitual' and performed rather automatically and without conscious deliberation (Wood et al., 2002; Wood & Neal, 2009).

Despite the routine-intensive nature of many jobs from cashiers to surgeons, and traditional office work[1], prior research on habit formation has largely focused on non-work contexts, such as physical fitness, personal finance, and public transport usage (Allcott & Rogers, 2014; Carrera et al., 2020; Charness & Gneezy, 2009; Giuntella et al., 2024; Gravert & Collentine, 2021; Royer et al., 2015; Yang & Long Lim, 2018), domains that are characterized by different motivational structures compared to organizational settings.

In workplace settings, the effectiveness of temporary incentives in fostering long-term habits remains unclear. As summarized in the review by Gneezy et al. (2011), temporary incentives can help establish lasting habits in areas such as health, where individuals may come to appreciate the intrinsic benefits, such as improved well-being after initially engaging in the behavior due to external rewards. In work contexts, however, such intrinsic benefits may be less apparent to employees. Moreover, contract theory would predict that temporary incentives raise performance only temporarily, i.e. only as long as they lead to higher rewards but not afterwards. Beyond that, there is a risk that temporary incentives could lead to crowding-out effects, ultimately reducing employees' intrinsic motivation for the respective task below its original level (Deci, 1971; Falk & Kosfeld, 2006; Gneezy & Rustichini, 2000a; Lepper et al., 1973). This effect may occur for example if the incentive inadvertently signals that the task is undesirable (e.g., 'the task is very difficult') or if it conveys unfavorable social norms (e.g., 'nobody else does it') (Alfitian et al., 2024; Bénabou & Tirole, 2003; Campos-Mercade et al., 2025; Gneezy et al., 2011; Sliwka, 2007).

[1]Note that such routine tasks "require methodical repetition of an unwavering procedure" (see Autor et al., 2003, p. 1283) and it has been argued that this very repetition is a key ingredient in habit formation as it allows behaviors to become automatic over time (Wood & Neal, 2009).

In this paper, we investigate whether and how firms can foster the formation of habits at work. We follow a behavioral definition of habit formation, whereby habits are regularly repeated behaviors that are *insensitive to changes in the reward structure* (Dickinson, 1985; Gillan et al., 2015). Therefore, if employees start engaging in a new behavior due to an extrinsic reward, but continue to do so after the incentive is removed, this indicates that a habit has formed.[2] We study this question in a field experiment with 829 employees in 15 technical service firms. These firms provide on-site installation and maintenance services on behalf of a large multinational company – hereinafter referred to as *MultiCo*. To support sales efforts, MultiCo seeks to encourage its technicians to leverage their on-site presence as an opportunity to advise clients on new or additional products and services. However, although the company has previously attempted to increase sales lead generation, these efforts have been largely unsuccessful, potentially due to the highly technical nature of the technicians' vocational training and primary job roles. To understand how firms can support habit formation in the workplace, we randomized the implementation of a temporary monetary incentive that rewards sales-oriented behavior performed on a regular basis. Technicians in the treatment group received a €100 bonus for generating at least one sales lead per week over a period of four consecutive weeks. Sales leads are potential or existing customers who have given their consent to be contacted regarding the purchase of a specific product. The incentive was offered for a fixed period of 12 weeks, which was clearly communicated upfront to employees in the treatment group, allowing them to earn up to €300 in total.

We investigate two key pre-registered hypotheses: (i) the temporary incentive increases sales activities during its implementation, and (ii) this effect persist after the incentive has been discontinued. We test different cognitive mechanisms for the second hypothesis to better understand the formation of work habits and to provide guidance for promoting them. For one, a behavioral shift may be due to increased *automaticity*. That is, while prior to the intervention many employees were not used to regularly talk to customers about potential sales, they now acquire the habit to do so. As argued in a rich literature in psychology (Verplanken & Orbell, 2022; Wood et al., 2002; Wood & Rünger, 2016), the fact that an activity is repeated frequently can lead to a certain automaticity of the behavior, i.e. an action is performed unconsciously and "out of habit" rather than through deliberate choice. Moreover, as lined out by Chapman and Gneezy (2024), temporary

[2]Other definitions include additional assumptions on the cognitive foundations of habits, such as cue-response mechanisms (Beshears et al., 2021; Mazar & Wood, 2018; Wood & Rünger, 2016) or (un-)conscious behaviors (Verplanken & Orbell, 2022). See Volpp & Loewenstein (2020) and Chapman & Gneezy (2024) for discussions regarding different definitions of habit formation.

incentives can also lead to a permanent shift in behavior due to *taste acquisition*: By repeatedly engaging in a new task, people may develop "a taste" for the new behavior, as they overcome initial aversion to certain aspects of the task. A related argument had already been put forward in Stigler and Becker (1977) and Becker and Murphy (1988) who note that repeated engagement in a task can build what they term "consumption capital": a stock of experience that makes an activity more rewarding over time. In our context, as the temporary incentive motivates technicians to perform sales activities more often, they may learn to enjoy customer interactions and continue to do so after the incentive has been discontinued. Another related channel is *human capital formation*: By repeatedly performing a task, employees learn from their experience, which reduces the costs of effort or increase productivity in the future (Arrow, 1962; Becker, 1965; Jovanovic & Nyarko, 1996). In comparison to the prior explanation, the employees' motivation for the task itself remains unchanged, but the task becomes easier to perform due to larger task-specific human capital. Finally, the temporary incentive can also provide *direction*, as it may credibly reveal that their employer considers sales activities to be a promotable task or an important aspect of the overall strategy of the firm. This, in turn, may encourage employees to engage more persistently in these activities, for instance due to career concerns (Holmström, 1999, 2017).

In line with our pre-registered hypotheses, we find a significant increase in sales leads not only during but also after the incentive phase. During the incentive period, the effect amounts to 31 percent of the pre-treatment mean. Although the effect after the incentive period is smaller in magnitude, it remains sizable at 15.5 percent higher performance relative to the the pre-treatment outcome. Successful sales leads – that is, leads resulting in actual sales – also increase by 16.8 percent during the incentive phase and this effect persists during the post-incentive phase at 13.3 percent more successful leads. Using a post-experimental survey we compare the potential mechanisms of habit formation laid out in the above. We find consistent evidence for the role of acquired taste: Technicians in the treatment group report a significantly higher level of intrinsic motivation for sales activities after the incentive has been discontinued compared to those in the control group. Moreover, we observe a similar pattern for another customer-oriented task. It thus appears that the temporary incentive indeed had persistent performance effects because it led technicians to "learn to like" customer interactions and, in turn, achieve more sales. Further analyses show that the increase in sales activities did not reduce customer satisfaction. Despite the large upfront bonus payments for generating sales leads during the incentive phase, a back-of-the-envelope calculation suggests that the intervention led to a profit increase already within a few months after the end of the incentive phase.

Our contribution to the literature is twofold. First, we complement the rich literature on incentives in organizations.[3] By showing that even temporary monetary incentives can be used to permanently align individual preferences and company goals, we underline the power of habits for organizational functioning. Previous studies on interventions targeting habit formation have focused on non-work behaviors such as exercise, commuting and health-related behaviors (Acland & Levy, 2015; Allcott & Rogers, 2014; Beshears et al., 2021; Celhay et al., 2019; Charness & Gneezy, 2009; Gertler et al., 2018; Giuntella et al., 2024; Larcom et al., 2017; Royer et al., 2015; Yang & Long Lim, 2018). Second, our study also contributes to the debate on whether and where monetary incentives can undermine intrinsic motivation (Bénabou & Tirole, 2003; Frey & Oberholzer-Gee, 1997; Gneezy & Rustichini, 2000a, 2000b; Sliwka, 2007). To the best of our knowledge, our study is the first to provide (field) experimental evidence showing that monetary incentives can foster taste acquisition at the workplace (Chapman & Gneezy, 2024; Loewenstein & Angner, 2003) and thus *increase* rather than decrease intrinsic motivation after the incentive has been removed.

The remainder is structured as follows. Section 3.2 details the theoretical background. Section 3.4 and Section 3.3 describe the setting and the experimental design. Section 3.5 provides results on the temporary incentive effect and the underlying mechanism. Section 3.6 presents additional results on customer satisfaction and profit. Section 3.7 concludes.

[3] For surveys, see e.g. Prendergast (1999), Bandiera et al. (2011), Lazear (2018), and Mahlendorf and Vogelsang (2024) .

3.2 Theoretical Background

As firms adjust their strategies in response to evolving technologies or market conditions, employees need to adapt to changing processes and acquire new skills. Such transformations present managers with new challenges, as they seek to provide guidance regarding expected behaviors and, in particular, the motivation to embrace new tasks and processes. One way to address these challenges is through monetary incentives. It has often been argued that monetary incentives can effectively be used to align employee actions with company objectives. For instance, incentives can motivate employees by rewarding task completion (Eisenhardt, 1989; Holmström, 1979; Ross, 1973) or provide direction by highlighting priority tasks (Manthei et al., 2023). Numerous empirical studies have indeed shown that monetary incentives can raise performance during the period they are made available (Casas-Arce & Martinez-Jerez, 2009; Castro et al., 2025; Friebel et al., 2017; e.g. Lazear, 2000). In line with these findings, we expect that the temporary incentive leads to an upward shift in performance in the incentivized task during the time it is in place.

H1. The monetary incentive increases performance in the incentivized task during the incentive period.

Although the short-run effectiveness of monetary incentives is rarely disputed, it is not clear whether the long-term effect of temporary incentives is positive, neutral, or even negative. A key concern is crowding-out effects (Deci, 1971; Falk & Kosfeld, 2006; Gneezy & Rustichini, 2000a; Lepper et al., 1973), where removing incentives lowers engagement below pre-incentive levels, for instance, due to reduced intrinsic motivation or detrimental effects on norms of behavior.[4] In a workplace context, Alfitian et al. (2024), for instance, have recently shown that an incentive for perfect attendance backfired because it shifted social norms, making absenteeism more acceptable. However, for a crowding-out effect to occur, there must be a high level of intrinsic motivation or a strong social norm associated with the task at the outset. For novel or newly prioritized

[4] In their review on the long-term effects of monetary incentives, Gneezy et al. (2011) discuss two pathways towards a crowding-out effect: First, incentives can change the perceived locus of causality, i.e., an employee's intrinsic motivation to engage in a task is replaced by the external incentive (Brink et al., 2013; e.g. De Charms, 1968; Huffman & Bognanno, 2018). The second class of mechanisms encompasses signaling effects, where monetary incentives signal to the employee that a task is more difficult or less desirable than previously thought (Bénabou & Tirole, 2003; Cardinaels & Yin, 2015; Danilov & Sliwka, 2017; Sliwka, 2007). Similarly, (high) incentives can further cause repugnance, if they are perceived to be coercive or unethical (Ambuehl et al., 2015; Campos-Mercade et al., 2025; Stüber, 2024).

tasks this appears unlikely. Instead, we argue that for such tasks providing temporary incentives may help employees to align their attitude towards the behavior desired by the employer, as we will explain in the following.

Generally speaking, if a behavior persists after the initial incentive has been removed, this is often broadly defined as *habit formation* (Dickinson, 1985). The underlying processes behind this behavioral pattern can be very diverse. To organize potential mechanisms for habit formation (see Volpp and Loewenstein (2020) and Chapman and Gneezy (2024) for recent reviews), we classify them according to the amount of cognitive deliberation that is involved in the action. On the end points of this continuum, we consider actions that are either *automatic* – which describes actions that are effortless, intuitive or even subconscious – or *deliberate* – which involves strategic and effortful thinking (see Figure 3.1).[5]

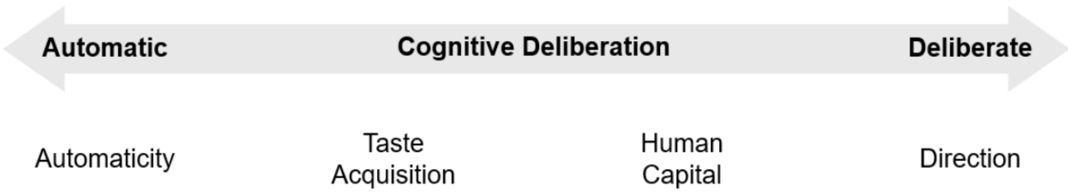

Figure 3.1: Potential Mechanisms for Habit Formation

Automaticity: A key component of many psychological definitions of habit formation is *automaticity* (Verplanken & Orbell, 2003; Wood et al., 2002; Wood & Rünger, 2016). According to Mazar and Wood (2018), automatic habits are "goal-independent, unconscious, efficient, fast, and stimulus-driven" (p. 92). Behaviors that are often cited in these contexts are exercise- and health-related (Buyalskaya et al., 2023; Charness & Gneezy, 2009; Gallani, 2023; Royer et al., 2015). Even if there was an initial conscious rationale behind some behaviors, they are often later on part of a subconscious routine that requires little to no reflection to be executed. Applied to the setting we study, consider technicians who receive multiple work assignments per day that require them to install or troubleshoot specific technical equipment at the customer's home. Given that the technicians have repeated these processes hundreds of times, they have a standard routine to

[5] Our classification maps well with the causes for management control problems described in Merchant and van der Stede (2017) An automatic response (automaticity) corresponds to avoiding the control problem through automation. The remaining mechanisms correspond to each of the causes of the control problem, i.e. motivational problems (taste acquisition), personal limitations (human capital) and a lack of direction (direction).

follow for every work assignment. If these technicians now receive an incentive to also ask customers whether they may be interested in additional products and services, they will likely initially do so to receive the bonus. However, over time this behavior may become part of their work routine. In this case, the technician at one point may no longer consciously engage with the customer in the hopes of receiving additional monetary benefits, but merely because it is now part of their unconscious routine.

Taste Acquisition: Following Chapman and Gneezy (2024), habits may also be formed as a result of *taste acquisition*. That is, continuous repetition of a behavior may lead a person to change their preferences towards a certain behavior in a positive way. As a result, their preferences are better aligned with their actions and the behavior is more intrinsically rewarding.[6] In contrast to automaticity, taste acquisition is a more deliberate process, because it often requires overcoming a certain level of personal resistance. However, once a preference shift has been realized, the activity is pursued deliberately. Returning to our example, when a high monetary incentive leads the technicians to engage with customers more frequently, it may enable them to discover the positive side of customer interaction (Loewenstein & Angner, 2003; Melchionne, 2007). This mechanism closely mirrors the economic view that repeated engagement can build what Stigler and Becker (1977) and Becker and Murphy (1988) describe as *consumption capital*: a stock of experience that makes an activity more rewarding over time. As noted by Stigler and Becker(1977), a similar idea already appears in Marshall's Principles of Economics, where he acknowledges in a qualification to the law of diminishing marginal utility "*that the more good music a man hears, the stronger is his taste for it likely to become*" (see Marshall, 1920, p. 79).

Human Capital: While taste acquisition is focused on people changing their motivational stance towards a particular activity, it is further possible that the incentive leads to an increase in human capital either through stronger incentives to acquire knowledge or through learning-by-doing. As employees refine their skills through continuous practice, it becomes easier for them to engage in the behavior in the future (Arrow, 1962; Becker,

[6] One might argue that acquiring tastes is conceptually similar to the reduction of cognitive dissonance that occurs whenever someone engages in an undesirable activity for the sake of a desirable outcome. (Festinger, 1957; Harmon-Jones & Mills, 2019). According to the theory of cognitive dissonance, employees might rationalize their sales activities by re-framing them as more useful for the customer or by selectively remembering pleasant customer interactions. However, the theory further predicts that dissonance reduction is substantially reduced when the rewards for the unpleasant activity are sufficiently large (Festinger & Carlsmith, 1959). Additionally, the theory provides no explanation for long-term effects once the incentive has been removed. In case of a temporary incentive, dissonance reduction is no longer necessary once the incentive has been removed, and employees would therefore discontinue their sales activities in response to that.

1965; Jovanovic & Nyarko, 1996). In this case, the long-term effect of the incentive is driven by an increase in knowledge and skills, while the underlying motivation remains unchanged. For example, when technicians receive monetary incentives for generating sales leads, they may be more motivated to engage with learning materials or ask colleagues for advice. Similarly, as they gain experience in customer service, their conversational skills improve, leading to higher customer satisfaction ratings and - in turn - more sales. However crucially, if sales activities were no longer lucrative or became increasingly difficult, employees would scale back their efforts accordingly.

Direction: Finally, it is possible that the temporary incentive system sends a signal about the firms priorities or overall strategy (Castro et al., 2025; Sandvik et al., 2021). In the spirit of Holmström: "Putting money behind a measure conveys a stronger message of what is expected" (2017, p. 1772). In that sense, technicians may infer from the bonus for sales activities that engaging in sales behavior is of high importance for the firm. In turn, such behavior can then appear more relevant for career progression and wage increases, in line with the classical model of career concerns (Holmström, 1999).[7] In summary, we argue that temporary incentives can lead to habit formation in the workplace and we test each of the presented mechanisms and alternative explanations for a persistent behavioral shift using data from a post-experimental survey.

H2. The monetary incentive increases performance in the incentivized task after the incentive has been discontinued.

3.3 Institutional Setting

Our sample consists of all active technicians from 15 medium-sized technical service firms who install and maintain products and services on behalf of MultiCo. MultiCo both employs its own technicians and works with these service firms as sub-contractors. MultiCo's field of work is divided into 37 regions. Individual employees at the subcontractors are typically assigned to specific local areas within a region and receive their work schedule every morning from their local dispatcher.[8] While some assignments

[7]Similarly, the incentive may cause what the accounting literature terms strategy surrogation, setting a strong focus on the specific performance measure used (Choi et al., 2012, 2013; Wang et al., 2023).

[8]Dispatchers were not made aware of the experiment and generally do not have much influence on the distribution of work assignments across technicians themselves, as they receive a pre-selected 'basket' of assignments from the MultiCo dispatch. It is therefore unlikely that the dispatchers deliberately allocated more personal visits to technicians in the treatment group to create more sales opportunities.

can be performed remotely, many assignments require a personal visit at the customer's home. Even though MultiCo's employees operate in the same infrastructure, interaction between employees, even within the same team members is very limited.

To leverage technicians' customer contact, MultiCo has increasingly instructed technicians to use personal visits as an opportunity to promote and sell additional products and services. While a few products can be provided directly from the technician's mobile inventory, the larger volume is processed by MultiCo's customer service. If a customer is interested in additional products or services, employees create a so-called *sales lead* that is handled by MultiCo's customer service. If the customer service agent manages to sell an additional product to the customer, the lead is considered *successful* and the technician who created the lead receives a commission for the successful sale. Employees have access to a list that shows which products are incentivized and how large the commission for a specific product is. Products with a higher profit margin also lead to higher commissions.[9] As the processing of generated sales leads and actual sales do not occur immediately after the customer visit, there are often substantial time lags until a bonus payout occurs.[10] To support the service employees in identifying sales opportunities and approaching the customers, MultiCo offered online sales trainings to all employees before and at the time of the experiment. Training sessions took place outside of regular working hours.[11] However, neither the existing bonus scheme for successful sales nor the online trainings led to a substantial increase in technician's sales activities.

3.4 Experimental Design and Data

We randomize the implementation of an additional temporary bonus scheme to foster regular sales activities among technicians. Employees in the treatment group receive a bonus of €100 for generating at least one sales lead per week for four consecutive weeks. Importantly, different from the existing sales commission scheme (which remains in place), the sales lead in this new temporary bonus scheme does not have to be successful for employees to receive the bonus. The intervention runs for twelve weeks, such that

[9] Commissions can range from less than €10 to up to €175 per product sold. A single sales lead can result in the sale of more than one product. On average, 38.71% of sales leads during the pre-period are successful. The average commission earned from each successful sales lead for an incentivized product is €22 in that time frame.

[10] Commissions for successful leads prior to the experiment are typically paid out within eight weeks, with 75% issued during that time frame. However, certain products and services have longer wait times due to their complexity or the need for extensive paperwork.

[11] As only few employees took up the offer, the training sessions were temporarily discontinued around one month after the start of the experiment.

employees can receive the bonus up to three times, resulting in a bonus of up to €300.[12] After the end of the twelve week intervention period, we observe individuals for a twelve week post-intervention period in order to assess if there is a persistent behavioral change regarding the reporting of sales leads.[13]

The experiment started in March 2023. Employees were informed up to 1.5 weeks prior to the experiment.[14] We randomly assigned 923 individual employees (from the total of 15 firms) to treatment and control groups using stratified randomization based on the employee's company and their own previously reported sales leads (dummy equal to one if they would have received at least one bonus in the last 12 full weeks and zero otherwise, and a dummy equal to one if they reported sales leads since October 2022 and zero otherwise). Hence, we randomize treatment assignment *within* firms to be unaffected by potential differences in firm-specific time trends. To assure that we consider only employees who benefit from the bonus, we restrict our primary analyses to the 829 employees who remained in the firm until the end of the post-intervention period.[15]

As Table 3.1 shows, the sample is well balanced with respect to our main outcomes and across firms. Service technicians in the incentive group as well as the control group report around 1.3 sales leads per week on average in the pre-period, 0.5 of which were successful on average. Also, the regularity with which sales leads are reported is not statistically significantly different between groups. In both groups, the average employee would have received about 0.5 bonuses in the pre-period if the scheme had been in place already at this point in time.

[12]A service technician receives a gross wage of around €2,900 per month. Thus, they can get around 3.4% extra through the incentive.

[13]For fairness reasons so that no employee is disadvantaged by the random assignment, technicians in the control group also receive the new bonus scheme for twelve weeks. This period started two weeks after the end of the post-intervention period.

[14]See Figure 3.4 to Figure 3.6 in the Appendix for the wording of the email announcements for the respective groups.

[15]As Table 3.6 in the Appendix shows, there is no evidence for selective attrition based on time or prior sales performance. The table shows that 9.8% of service technicians in the incentive group and 10.6% in den control group left their firm until the end of the post-intervention period. During the incentive phase, 4.6% (incentive group) and 4.8% (control group) left their employer. In the first four weeks after the incentive phase, 1.3% and 0.6% left (p-value=0.313).

Table 3.1: Balance Check

	Incentive Group		*Control Group*		
	Mean	S.D.	Mean	S.D.	p-value
Weekly Leads Pre	1.271	1.741	1.333	1.981	0.637
Weekly Succ. Leads Pre	0.503	0.751	0.504	0.729	0.981
Bonuses Pre	0.493	0.694	0.525	0.712	0.504
Company 1	0.053	0.224	0.048	0.215	0.770
Company 2	0.075	0.263	0.080	0.271	0.772
Company 3	0.022	0.146	0.027	0.161	0.639
Company 4	0.149	0.357	0.128	0.335	0.388
Company 5	0.053	0.224	0.053	0.225	0.980
Company 6	0.082	0.274	0.094	0.293	0.519
Company 7	0.137	0.344	0.140	0.348	0.887
Company 8	0.031	0.174	0.029	0.168	0.853
Company 9	0.058	0.233	0.063	0.243	0.750
Company 10	0.050	0.219	0.051	0.220	0.981
Company 11	0.058	0.233	0.051	0.220	0.664
Company 12	0.123	0.328	0.126	0.332	0.885
Company 13	0.038	0.193	0.036	0.187	0.871
Company 14	0.022	0.146	0.024	0.154	0.804
Company 15	0.050	0.219	0.048	0.215	0.891
Observations	416		413		829

Note: This table reports summary statistics separately for treatment and control group. Additionally, we report p-values for t-tests comparing the means of the continuous variables and test of proportions for the dummy variables, respectively. The table is based on out balanced sample, i.e. only including individuals who are still there at the end of our observation period.

3.5 Results

3.5.1 Effect on Sales Leads

To investigate the effect of our bonus intervention on the technicians' behavior, we first examine the number of reported sales leads over the span of our observation period. Figure 3.2 plots the number of sales leads by treatment group over the weeks since the incentive start (first week with incentive is week 1). The graph starts with the beginning of 2023. The red line at week 0 marks the last week before the start of the incentive phase. The employees were already informed about the upcoming bonus during week -1. The red line in week 12 marks the last week of the incentive phase. We use the balanced panel, i.e. only include individuals who are still there at the end of our observation period.

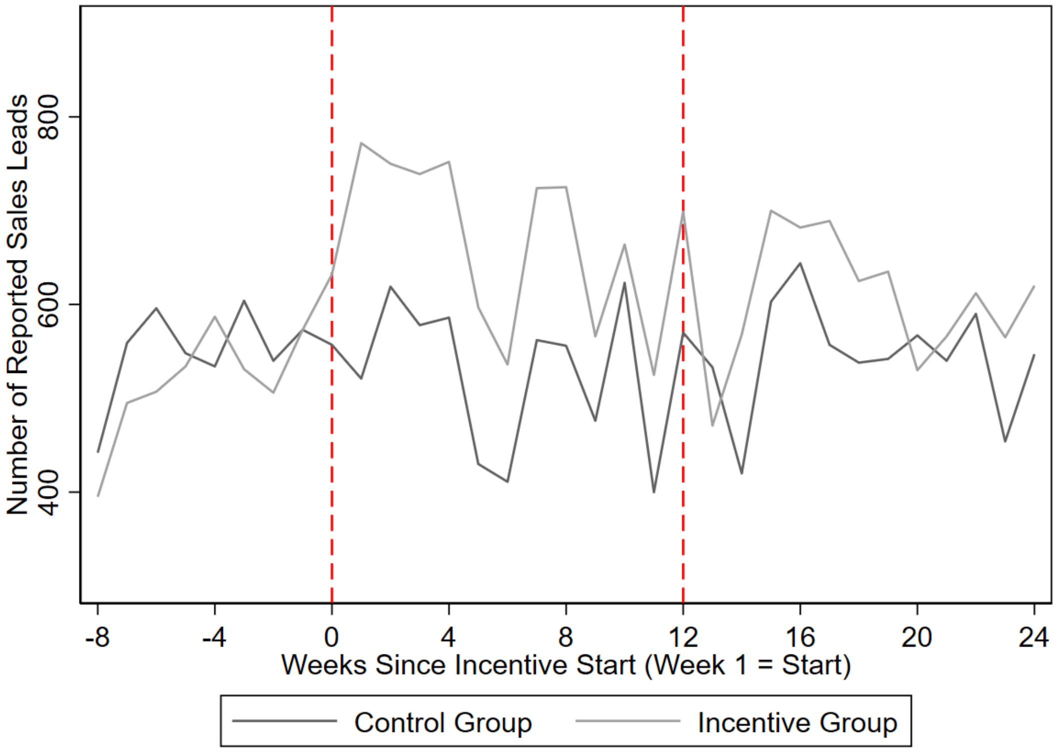

Figure 3.2: Sales Leads over Time

Table 3.2: Effect on the Number of (Successful) Sales Leads

	Sales Leads$_{\mathrm{it}}$ (1)	*Successful Leads*$_{\mathrm{it}}$ (2)
$Treat_{\mathrm{i}} \times Incentive_{\mathrm{t}}$	0.396***	0.084**
	(0.073)	(0.035)
$Treat_{\mathrm{i}} \times Post\ Incentive_{\mathrm{t}}$	0.197**	0.067*
	(0.100)	(0.038)
p-value Inc=Post	0.014	0.569
Time Fixed Effects	Yes	Yes
Individual Fixed Effects	Yes	Yes
Clustered at	Individual	Individual
Number of Clusters	829	829
Observations	27,357	27,357
Adjusted R-squared	0.565	0.409

Note: This table reports results of a difference-in-differences regression of the number of reported sales leads (column 1) or successful sales leads (column 2) on being part of the treatment group, i.e. incentive group, during as well as after the incentive phase in comparison to the pre-incentive phase. The data is on the individual-week level. The regression is based on the balanced panel, i.e. only include individuals who are still active for the firm at the end of our observation period. We include individual as well as week fixed effect. Standard errors are clustered at the individual level, and reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

As Figure 3.2 shows, here is a rather sharp increase in reported sales leads in the treatment group when the treatment is implemented[16] and the number of reported leads stays larger in the treatment than in the control group during the 12 week treatment phase. But the number of reported leads also tends to be higher on average even after the end of this period.

As the technicians received the bonus for reported sales leads independent of the success of these sales leads, it is conceivable that technicians simply increase the number of reported sales leads even when these customers do not have an intention to buy further products or services. However, as column (2) in Table 3.2 shows, the treatment also increased the number of successful sales leads per technician and week on average by 0.084, which is around 16.8% in comparison to the pre-period mean of 0.5 successful sales leads. Again the coefficient is somewhat smaller for the post-intervention period

[16]Recall that the treatment was announced about 1.5 weeks before the bonus was active which started to be payoff relevant in week 1. There seems to be an anticipatory effect already in week 0 which may be the result of technicians already trying out how to approach customers about additional services before the actual start of the incentive period.

but technicians reported sales leads still lead to 0.067 or about 13% more successful sales even after the monetary incentive has expired.[17] Thus, while not each additional sales lead generated through the incentive is successful, the incentive also translates into a sizable increase in successful sales leads during and after the incentive phase.[18]

We further investigated whether the incentive for regularly reporting sales leads not only increased the number of (successful) sales leads but also their regularity. To test that, we also consider the number of bonuses a technician would have received (i.e. when all had been part of the bonus group all the time; that is the number of consecutive four week periods with a sales lead in each) as well as the number of weeks with at least one sales lead as outcome variables. Table 3.9 in the Appendix shows the results. While the number of bonuses a technician would have received only significantly increases during the incentive period, the number of weeks with at least one sales lead increases both during the incentive as well as the post-intervention period. Thus, even though technicians do not seem to try anymore to reach exactly four weeks in a row with sales leads, they still create sales leads more regularly. Overall, we find evidence for habit formation in response to temporary incentives that is persistent beyond the incentive period. Thus, our results support *H1* and *H2*.

3.5.2 Mechanisms

To understand the behavioral mechanisms underlying our findings, we conducted a survey during the post-intervention phase. The survey contains a set of items for each of the potential behavioral channels described in Section 3.2, namely, *automaticity*, *taste acquisition*, *human capital formation*, or *direction*. All participants were assured at the beginning of the survey that individual answers were pseudonymized and would, under no circumstances, be shared with supervisors or MultiCo employees.

We used the automaticity subscale of the self-reported habit index (Verplanken & Orbell, 2003) as a measure for *automaticity*, which is a standard scale with high predictive validity that has been used in a multitude of studies such as Gardner et al. (2011, 2012, 2022).[19] To assess *taste acquisition* we use a common measure of *intrinsic motivation* for a

[17]The percentage of successful sales leads in the treatment group during the post-period is 38.84%. The average commission earned per successful sales lead for an incentivized product in this timeframe is €22.51. Thus, there is only a very slight change compared to the pre-period, which had 38.71% successful sales leads and an average commission of €22.

[18]Table 3.8 in the Appendix shows the results when including all service technicians we initially randomly assigned. Their number of sales leads is reported as 0 for the weeks in which they were no longer able to report sales leads.

[19]Items are: "Talking about additional services/products when visiting customers is something …" 1.

Table 3.3: Summary Statistics for Survey Responses

	Incentive Group		*Control Group*	
	Mean	S.D.	Mean	S.D.
Automaticity Index	4.749	1.839	4.541	1.768
Intrinsic Motivation	4.755	1.752	4.231	1.526
Knowledge	3.481	1.032	3.476	1.024
Perceived Importance	5.594	1.432	5.359	1.399

Note: This table reports summary statistics for the survey responses. Responses for items on automaticity, intrinsic motivation, and perceived importance are measured on Likert scales ranging from 1 (lowest) to 7 (highest). For each construct, we calculate the mean of all corresponding items. Knowledge is assessed as the number of correct answers in a short quiz consisting of five questions.

task using three items from the interest/enjoyment sub-scale of the Intrinsic Motivation Inventory (Ryan et al., 1983) adapted to the sales task.[20] In order to assess treatment effects on *human capital* acquisition, we conducted a short quiz to test the employees' knowledge about sales activities. All questions were based on existing training material of the firm that was regularly used to teach technicians how to effectively generate promising sales leads.[21] Finally, we measured the perceived importance of sales activities to determine whether the temporary incentives provided technicians with *direction* on work priorities expected of them. These questions capture, for instance, whether technicians perceive sales activities to have a high priority for managers or the firm in general.[22]

The survey was distributed to all service technicians from 12 out of 15 companies who were active at that time (4 weeks after the end of the incentive phase).[23] The invitation

"… I do automatically", 2. "… I do without having to consciously remember", 3. "… I do without thinking", 4. "… I start doing before I realize I'm doing it" (Scale: 1 (do not agree at all) - 7 (Fully agree); Cronbach's Alpha: $\alpha = 0.91$).

[20]Items are: "Talking about additional services/products when visiting customers is something …" 1. "… that gives me great pleasure", 2. "…that I enjoy", 3. (reversed) "… that I find boring" (Cronbach's Alpha: $\alpha = 0.70$).

[21]We measure that using five closed-ended questions regarding how to generate sales leads, e.g. "To make sure the customer doesn't feel pressured into making a purchase, I can use phrases like"Should we take a look?" or "Should we try this?". As a further validation we also repeatedly posted the questions (in the original German wording) to ChatGPT-4o prompting it to consider best practices and scientific insights when answering. In all cases, the answers provided were identical to those deemed correct in the knowledge quiz. When asked to assign a certainty level to these answers, the confidence levels ranged from 85% to 100%.

[22]The items are: "Talking about additional services/products when visiting customers is something …" 1. "… that my boss expects of me", 2. "…that is one of the important tasks of my job", 3. "… that is demanded of me" (Cronbach's Alpha: $\alpha = 0.67$).

[23]Due to legal regulations, each firm had to grant the research team the authority to directly contact their employees. While 12 of them agreed, the remaining 3 firms did not allow us to send out the surveys. Given that the firms that declined their participation are among the smallest firms in the sample (we contacted

was adapted from previous MultiCo surveys and did not mention sales activities or incentives as a topic of interest. Note that the survey specifically asks about the technicians' experiences in the last two weeks, i.e., only in the post-treatment period. Participants were paid a participation fee (€8) conditional on completing the survey. The response rate was around 30%.[24] We do not find that the share of survey participation differs between treatment and control group (p-value = 0.754). Table 3.3 shows descriptive statistics for the four measures for the treatment and control group respectively.

Table 3.4 presents the regression results where the standardized survey constructs are regressed on the treatment dummy. We do not find significant effects of the treatment on perceived automaticity in behavior, the direction through the incentive that sales is an important task, or human capital acquisition. However, we find a significantly higher level of intrinsic motivation for sales activities in the incentive group. In line with recent ideas, such as those in Chapman and Gneezy (2024), the monetary incentives appear to have led employees to acquire a taste for the target behavior, which persisted even beyond the incentive period.

To test this *taste acquisition* channel further, we can make use of additional survey items about technicians' preferences for another customer-oriented task, namely calling customers before visiting them. Employees are strongly encouraged to contact the customer prior to the appointment to personally introduce themselves to the customer and to get a better understanding of their issue. We indeed observe a similar increase in intrinsic motivation for contacting customers before the visit.[25] Therefore, these findings suggest that temporary incentives even created spillover effects on other customer oriented tasks. This supports the view that many technicians initially viewed the sales task more as a nuisance. The monetary incentive then provided material incentives to perform the task and to seek out customer contact at a regular frequency. In turn, this seems to have led a sufficient number of technicians to realize that customer contact can be enjoyable.

While the treatment coefficients in the regressions for the *automaticity* and *direction* measures are statistically insignificant, the point estimates are not sufficiently small to rule out the relevance of these mechanisms altogether. But concerning the *direction* channel, it is important to note that the respective survey responses show that employees are well aware at the outset that sales has a strong priority: As Table 3.3 shows, even employees

92% of all technicians in our sample) and as treatment assignment was randomized within firms, we do not consider this a threat to validity.

[24]However, not all participants answered every item. See Table 3.10 in the Appendix for results based only on participants who answered every item.

[25]See Table 3.11 in the Appendix.

Table 3.4: Survey Results

	Automaticity$_i$	*Taste Acquisition*$_i$	*Human Capital*$_i$	*Direction*$_i$
	(1)	(2)	(3)	(4)
Treat$_i$	0.115	0.315**	0.005	0.166
	(0.127)	(0.126)	(0.126)	(0.126)
Observations	248	245	253	252
R-squared	0.003	0.025	0.000	0.007

Note: This table presents the results of a regression analysis examining the effect of being in the treatment group (i.e., the incentive group) on measures of Automaticity, Taste Acquisition, Human Capital and Direction. Automaticity is measured using the four items from the Automaticity sub-scale of the Self-Reported Habit Index in Verplanken and Orbell (2003). Taste acquisition is measured using three items from the Intrinsic Motivation Inventory by Ryan, Mims, and Koestner (1983). Human Capital is measured using a quiz consisting of four questions based on MultiCo's training material. Direction is measured using three items assessing the perceived importance of the task. Questions are answered on a scale from 1 = "fully disagree" to 7 = "fully agree". An exception is our measure for human capital, where the answers are always "yes" or "no". All measures are calculated as the means of all items, which are standardized using z-scores. The data is at the individual level. There are slight differences in the number of observations across the constructs, as not all survey participants answered every item. See Table A10 in the Appendix for results based only on participants who answered every item. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

in the control group rate the perceived importance of the sales task as rather high, with a mean of 5.4 on a scale from 1 (lowest) to 7 (highest). Thus, a lack of understanding of the task's importance appears unlikely to explain why employees do not engage more in sales activities in the first place.

3.6 Further Results

3.6.1 Customer Feedback

We conducted additional analyses on two further aspects: customer satisfaction and sales profit. Although we did not formulate specific hypotheses regarding these aspects, they are essential for evaluating the success of this bonus scheme. While our incentives increase the number of sales leads, they may also lead to dissatisfaction of customers in case they feel overly pressured into agreeing to talk about additional products and services. Hence, we analyze the effect of our incentive on customer satisfaction ratings of the technicians' visits. After the technicians complete their visit, customers receive an automated feedback request by MultiCo where they are asked to rate the technicians visit on a five-star rating scale ranging from 1 star (worst) to 5 stars (best).

Table 3.5: Effect on Customer Satisfaction Ratings

	Star Rating$_{\text{it}}$ (1)	*Number Lowest Rating*$_{\text{it}}$ (2)	*Number Top Rating*$_{\text{it}}$ (3)
$Treat_{\text{i}} \times Incentive_{\text{t}}$	0.040	-0.021*	-0.104
	(0.029)	(0.012)	(0.073)
$Treat_{\text{i}} \times Post\ Incentive_{\text{t}}$	0.014	-0.009	-0.108
	(0.028)	(0.014)	(0.083)
p-value Inc=Post	0.365	0.337	0.959
Controls	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes
Clustered at	Individual	Individual	Individual
Number of Clusters	783	783	783
Observations	19,693	19,693	19,693
Adjusted R-squared	0.290	0.148	0.310

Note: This table reports results of a difference-in-differences regression of the customer satisfaction rating (column (1)), number of lowest possible ratings (column (2)) or number of highest possible ratings (column (3)) on being part of the treatment group, i.e. incentive group, during as well as after the incentive phase in comparison to the pre-incentive phase. The data is on the individual-week level. The regression is based on the balanced panel, i.e. only include individuals who are still there at the end of our observation period. We control for the share of being on-time and fulfilling the task of the individual in the week. Furthermore, we include individual as well as week fixed effect. Standard errors are clustered at the individual level, and reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5 shows the results of a difference-in-differences regression of customer ratings on being in the incentive group during the incentive period or in the post-intervention period. We do not find a decrease in the average star rating for the visits. If anything, the coefficient is even positive albeit insignificant (column 1). The number of visits that received the lowest rating is even marginally significantly reduced during the incentive phase (column 2). However, also the coefficient for the number of five star ratings is negative but insignificant (column 3). Thus, overall we do not find evidence for a change in customer satisfaction due to the incentive scheme.

3.6.2 Profit

A natural question to assess the success of the incentive scheme is whether the increase in the number of successful sales leads can offset the additional costs for the bonuses and lead to higher profits. To explore this, we estimate the monthly net profits generated by a technician as total value added of the products sold based on sales leads reported by this technicians in this month (also if the actual sale was later in time) subtracting the total commission paid for the sales leads in that month. Profit margins are based on a simplified report we received from MultiCo.

Figure 3.3 plots the sales profit change in percent by treatment group over the calendar months of 2023. The profit is normalized such that the profit is the same for both groups in January (month 1). The red line at month 2 marks the last month before the start of the incentive phase. The red line in month 5 marks the last month of the incentive phase. We use the balanced panel, i.e. only include individuals who are still there at the end of our observation period. As the figure shows, the percentage change in profits relative to the January profits separately for incentive and control group over time. As the Figure shows, the incentive group generates lower profits during the incentive phase as the costs of the incentive scheme outweigh the benefits during this phase. In the post-incentive period, however, the incentive group tends to generate larger profits as technicians in the treatment group continue to provide more successful sales leads as already shown above, but the firm no longer has to pay the additional bonus payments. The corresponding regression results are shown in Table 3.12 in the Appendix. By a simple back-of-the-envelope calculation the break-even point is attained in month 8. As the control group received bonuses starting at the beginning of September (month 9), we cannot observe how long the positive effect on profit lasts. However, as it appears very unlikely that crowding-out effects occur with a strong time lag the temporary bonus is very likely to be profitable overall.

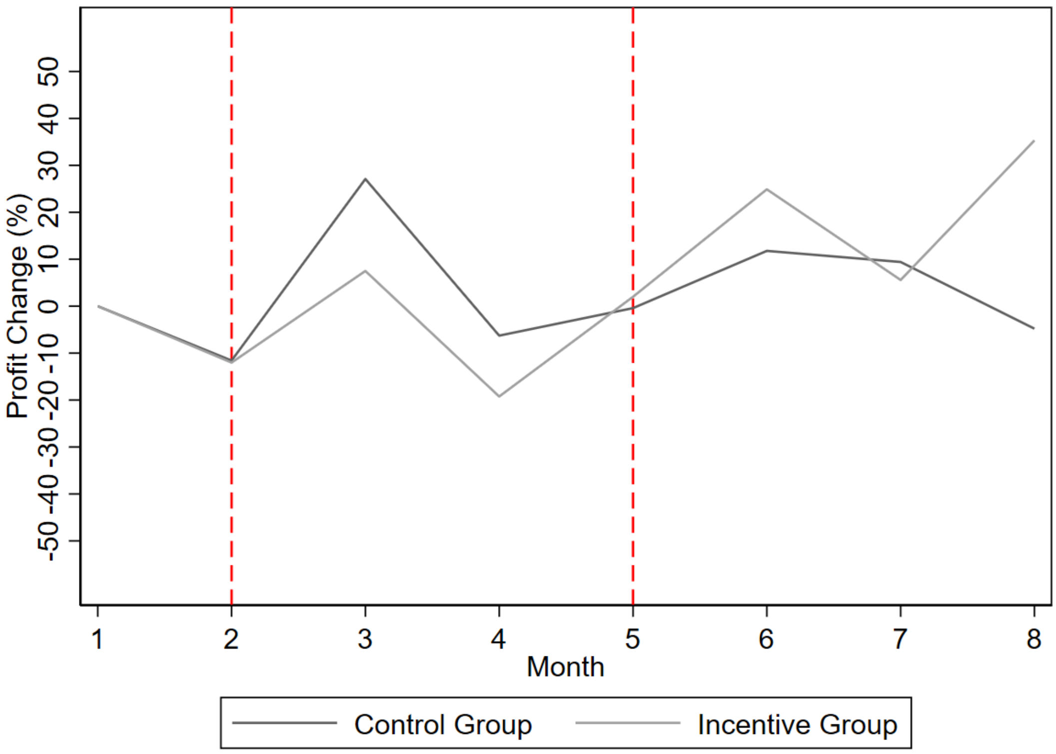

Figure 3.3: Sales Profit over Time

3.7 Conclusion

We present evidence showing that workplace habits can be shaped through temporary monetary incentives. In our setting, employees continued to perform a task at a higher intensity after the temporary incentive was removed. Thus, our findings complement prior research on the long-term effects of temporary incentives and management controls in general. While the previous literature has stressed the danger that monetary rewards may crowd out intrinsic motivation, we show that incentives can also lead to a "crowding-in" of task motivation at the workplace. We furthermore provide evidence supporting the role of taste acquisition (see Chapman & Gneezy, 2024) in this process: when monetary incentives motivate employees to perform a task more frequently, they may learn to like it and then perform it without further external reinforcement.

It is important to note that these results do not contradict previous findings on potential detrimental effects of temporary incentives. The results rather suggest that one should consider the prior task motivation or existing social norms. As laid out in the above, previous studies documenting detrimental effects of incentives typically investigated settings in which intrinsic motivation was rather high at the outset (Deci, 1971; Gneezy & Rustichini, 2000a, 2000b) or there were strong prior favorable social norms (Alfitian et al., 2024). When this is the case, detrimental effects indeed appear more likely to occur as there is more scope to reduce the level of intrinsic motivation or damage social norms of behavior. Our results show, however, that for tasks with relatively low levels of prior engagement the opposite can hold. While we could not directly compare the effects of the temporary scheme with a permanent bonus for sales leads, our back-of-the-envelope calculations suggest that the temporary incentive may be more profitable than a permanent scheme when leading to persistent changes of behavior without coming with additional costs in the long-term.

Our results have several implications for managerial practice. For one, we provide evidence that habit formation, which often has been discussed in the context of consumption habits or health-related behaviors, also matters in the workplace. Once employees have acquired a habit they may continue to follow it even without further external impulses. While standard incentive theory typically suggests that it is crucial to use stable bonus plans that pay employees based on value created, our results thus also imply that firms may well consider changing the incentivized objectives from time to time targeting very specific tasks, thereby guiding employees to adopt novel routines supporting strategic objectives.[26]

Taken together, even short-term incentive programs can have a lasting impact by fostering productive habits that persists beyond the incentive period. Such initiatives can help organizations align employee behavior with strategic goals, equipping managers during periods of transition with an additional tool to navigate evolving circumstances.

[26]See Manthei et al. (2021) for a related argument in the context of learning-by-doing.

Table 3.6: Check for Selective Attrition

	Incentive Group		*Control Group*		
	Mean	S.D.	Mean	S.D.	p-value
Attrition	0.098	0.297	0.106	0.308	0.671
Attrition Incentive Phase	0.046	0.209	0.048	0.213	0.882
Attrition 4 Weeks After	0.013	0.113	0.006	0.080	0.313
Weekly Leads Pre (Left)	0.886	1.504	0.907	1.646	0.950
Bonuses Pre (Left)	0.311	0.633	0.367	0.698	0.683
Observations	461		462		923

Note: This table reports summary statistics for the individuals who left during our observation period separately for the incentive and the control group. We also report p-values for tests on the equality of proportions for the attrition shares as well as for t-tests for the continuous variables. Weekly leads in the pre-period as well as the number of bonuses individuals could have earned in the pre-period are only compared for the subsample of individuals that left during our observation period.

3.8 Appendix

3.8.1 Tables

Table 3.7: Effect on Number of (Successful) Sales Leads (Cross-Sectional)

	Incentive Period		Post-Incentive Period	
	Sales Leads$_{\mathrm{i}}$ (1)	*Successful Leads*$_{\mathrm{i}}$ (2)	*Sales Leads*$_{\mathrm{i}}$ (3)	*Successful Leads*$_{\mathrm{i}}$ (4)
Treat$_{\mathrm{i}}$	0.393***	0.086***	0.201**	0.071**
	(0.073)	(0.033)	(0.102)	(0.036)
Company Fixed Effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Observations	829	829	829	829
Adjusted R-squared	0.749	0.619	0.624	0.600

Note: This table reports results of a regression of the number of reported sales leads (columns (1) and (3)) or successful sales leads (columns (2) and (4)) on being part of the treatment group, i.e. incentive group, during (columns (1) and (2)) as well as after (columns (3) and (4)) the incentive period. The data is at the individual level. For each individual, sales leads and successful sales leads are measured as the mean over the weeks within the respective period. The regression is estimated on the balanced panel, i.e., all individuals who remain in the sample at the end of the observation period. We include company fixed effects and control for pre-treatment means of individual weekly sales leads and successful sales leads. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.8: Effect on Number of (Successful) Sales Leads (All Assigned Individuals)

	Sales Leads$_{\mathrm{it}}$ (1)	*Successful Leads*$_{\mathrm{it}}$ (2)
Treat$_{\mathrm{i}}$ × *Incentive*$_{\mathrm{t}}$	0.368***	0.085***
	(0.070)	(0.033)
Treat$_{\mathrm{i}}$ × *Post Incentive*$_{\mathrm{t}}$	0.150	0.062*
	(0.095)	(0.035)
p-value Inc=Post	0.004	0.403
Time Fixed Effects	Yes	Yes
Individual Fixed Effects	Yes	Yes
Clustered at	Individual	Individual
Number of Clusters	923	923
Observations	30,459	30,459
Adjusted R-squared	0.565	0.427

Note: This table reports results of a difference-in-differences regression of the number of reported sales leads (column (1)) or successful sales leads (column (2)) on being part of the treatment group, i.e. incentive group, during as well as after the incentive phase in comparison to the pre-incentive phase. The data is on the individual-week level. The regression is based on all individuals who were randomly assigned before the start of the incentive phase. We include individual as well as week fixed effect. Standard errors are clustered at the individual level, and reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.9: Effect on Number of Weeks with Leads

	Total Bonuses$_{\text{it}}$	*Weeks with Leads*$_{\text{it}}$
	(1)	(2)
$Treat_{\text{i}} \times Incentive_{\text{t}}$	0.225***	0.880***
	(0.060)	(0.185)
$Treat_{\text{i}} \times Post\ Incentive_{\text{t}}$	0.044	0.487***
	(0.058)	(0.187)
p-value Inc=Post	0.003	0.022
Phase Fixed Effects	Yes	Yes
Individual Fixed Effects	Yes	Yes
Clustered at	Individual	Individual
Number of Clusters	829	829
Observations	2,487	2,487
Adjusted R-squared	0.529	0.766

Note: This table reports results of a difference-in-differences regression of the number of bonuses an individual would have gotten (column 1) or number of weeks with at least one sales lead (column 2) on being part of the treatment group, i.e. incentive group, during as well as after the incentive phase in comparison to the pre-incentive phase. The data is on the individual-phase level. The regression is based on the balanced panel, i.e. all individuals who are still there at the end of the observation period. We include individual and phase fixed effects. Standard errors are clustered at the individual level, and reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.10: Survey Results (Subset of Participants Answering All Items)

	Automaticity		*Intrinsic*			*Importance*		*Communication*	
	Leads$_i$	*Calls*$_i$	*Leads*$_i$	*Calls*$_i$	*Knowl-edge*$_i$	*Leads*$_i$	*Calls*$_i$	*All*$_i$	*Leads*$_i$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treat$_i$	0.125	0.234	0.338**	0.458***	0.050	0.039	0.059	0.149	0.124
	(0.168)	(0.163)	(0.155)	(0.150)	(0.184)	(0.179)	(0.168)	(0.180)	(0.176)
Observations	128	128	128	128	128	128	128	128	128
R-squared	0.004	0.015	0.037	0.069	0.001	0.000	0.001	0.006	0.004

Note: This table presents the results of a regression analysis examining the effect of being in the treatment group (i.e., the incentive group) on various survey constructs. The constructs measured in the survey include automaticity, intrinsic motivation, knowledge, perceived task importance, and communication frequency. Each construct is measured by calculating the mean of multiple survey items and then standardized using z-scores. Columns (1), (3), and (6) refer to sales leads. Columns (2), (4), and (7) correspond to another customer-oriented measure, specifically the requested calls made before customer visits. Column (5) reports the results related to knowledge on how to generate sales leads. Column (8) refers to the frequency of communication with the supervisor, while Column (9) focuses on how often that communication was about sales. The data is at the individual level and includes only individuals who answered all survey items. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.11: Survey Results (With Other Task)

	Automaticity		*Intrinsic*			*Importance*		*Communication*	
	Leads$_{\mathrm{i}}$	*Calls*$_{\mathrm{i}}$	*Leads*$_{\mathrm{i}}$	*Calls*$_{\mathrm{i}}$	*Knowledge*$_{\mathrm{i}}$	*Leads*$_{\mathrm{i}}$	*Calls*$_{\mathrm{i}}$	*All*$_{\mathrm{i}}$	*Leads*$_{\mathrm{i}}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treat$_{\mathrm{i}}$	0.115	0.105	0.315**	0.328**	0.005	0.166	0.151	0.079	0.105
	(0.127)	(0.125)	(0.126)	(0.127)	(0.126)	(0.126)	(0.126)	(0.126)	(0.126)
Observations	248	253	245	243	253	252	253	253	253
R-squared	0.003	0.003	0.025	0.027	0.000	0.007	0.006	0.002	0.003

Note: This table presents the results of a regression analysis examining the effect of being in the treatment group (i.e., the incentive group) on various survey constructs. The constructs measured in the survey include automaticity, intrinsic motivation, knowledge, perceived task importance, and communication frequency. Each construct is measured by calculating the mean of multiple survey items and then standardized using z-scores. Columns (1), (3), and (6) refer to sales leads. Columns (2), (4), and (7) correspond to another customer-oriented measure, specifically the requested calls made before customer visits. Column (5) reports the results related to knowledge on how to generate sales leads. Column (8) refers to the frequency of communication with the supervisor, while Column (9) focuses on how often that communication was about sales. The data is at the individual level. There are slight differences in the number of observations across the constructs, as not all survey participants answered every item. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.12: Effect on Sales Profit

	Individual Sales Profit$_{\mathrm{it}}$	
	Gross	*Net (With Bonus)*
	(1)	(2)
$Treat_{\mathrm{i}} \times Incentive_{\mathrm{t}}$	14.527	-13.999
	(17.931)	(17.776)
$Treat_{\mathrm{i}} \times Post\ Incentive_{\mathrm{t}}$	23.421	23.421
	(20.420)	(20.420)
p-value Inc=Post	0.585	0.022
Time Fixed Effects	Yes	Yes
Fixed Effects	Individual	Individual
Clustered at	Individual	Individual
Number of Clusters	829	829
Observations	6,632	6,632
Adjusted R-squared	0.582	0.579

Note: This table reports results of a difference-in-differences regression of the individual employee gross sales profit (column (1)) or the net sales profit, i.e. subtracting bonus payments (column (2)) on being part of the treatment group, i.e. incentive group, during as well as after the incentive phase in comparison to the pre-incentive phase. The data is at the individual-month level. The regression is based on the balanced panel, i.e. only includes individuals who are still there at the end of our observation period. We include individual as well as month fixed effect. Standard errors are clustered at the individual level, and reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

3.8.2 Figures

Dear colleagues,

We want to inspire our customers with the best service again this year. To make it worthwhile for you to stay on track with your sales activities, *MultiCo* will soon start an initiative that targets lead generation. As not everyone can take part in the initiative at the same time a draw was held in advance to determine who can participate when.

Congratulations, you are in the first group!
We will launch a 12-week initiative for you starting Monday, 06.03.2023: Connect Four!

Those who regularly engage with customers and submit lead can receive up to €300 in additional bonuses.

How does "Connect Four" work?

If you manage to submit at least one sales lead per week for four consecutive weeks, you will receive an additional bonus of €100[1]. If the streak is interrupted, it must be started again from the beginning. Those who manage to complete multiple four-week streaks can look forward to an even higher bonus.

What is incentivized?

Every lead that is submitted with the customer's consent and recorded on the *platform name* counts. This includes, for example, *product name*, *product name*, *product name*, or *product name*. Unlike previous initiatives, leads are not tied to a successful sale. However, all current bonuses still apply – so it's doubly worthwhile for you to submit a promising lead.
You can find more details about the bonuses in the attached description.

Would you like to brush up on your knowledge on how to generate sales leads?

Then take a look at our online course on lead generation for field service technicians! Here you will not only learn how lead generation works on a technical level, but also how you can generate leads more easily and without abandoning your identity as a technician.

The course takes place every Tuesday from 17:00 - 18:30 and you can register for a day of your choice via this link

[link]

Figure 3.4: Wording of the Email Announcement (Treatment Group)

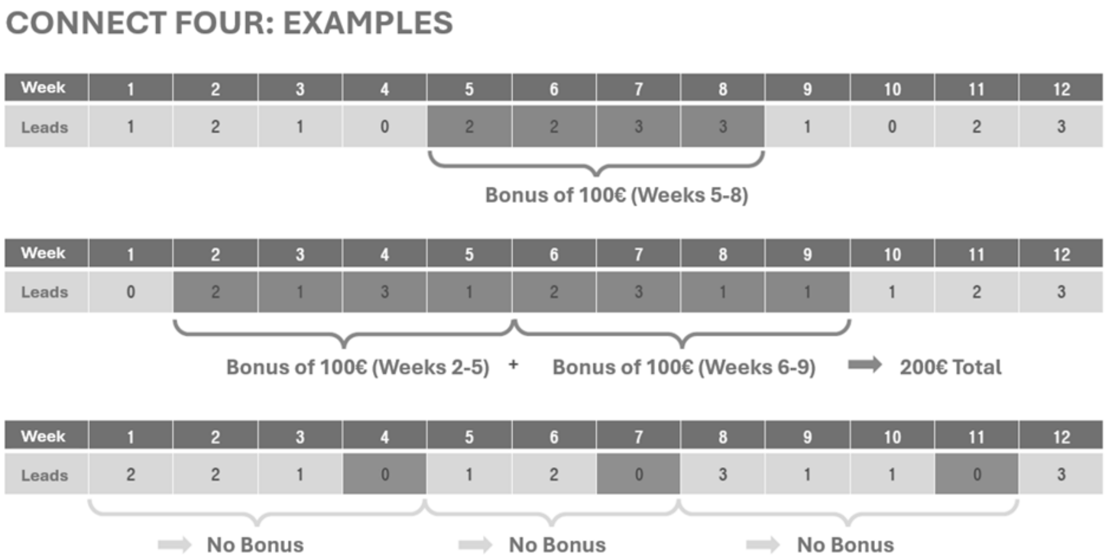

Figure 3.5: Email Appendix (Treatment Group)

Dear colleagues,

We want to inspire our customers with the best service again this year. To make it worthwhile for you to stay on track with your sales activities, *MultiCo* will soon start an initiative that targets lead generation. As not everyone can take part in the initiative at the same time, a draw was held in advance to determine who can participate when.

As soon as the initiative becomes relevant for you, you will be informed by us in good time!

Would you like to brush up on your knowledge on how to generate sales leads?

Then take a look at our online course on lead generation for field service technicians! Here you will not only learn how lead generation works on a technical level, but also how you can generate leads more easily and without abandoning your identity as a technician.

The course takes place every Tuesday from 17:00 - 18:30 and you can register for a day of your choice via this link:

[Link]

Figure 3.6: Wording of the Email Announcement (Control Group)

References

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In D. Card & O. Ashenfelter (Eds.), *Handbook of Labor Economics* (Vol. 4, pp. 1043–1171). Elsevier.

Acemoglu, D., & Pischke, J.-S. (1998). Why do firms train? Theory and evidence. *The Quarterly Journal of Economics*, *113*(1), 79–119.

Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in US wage inequality. *Econometrica*, *90*(5), 1973–2016.

Acland, D., & Levy, M. R. (2015). Naiveté, projection bias, and habit formation in gym attendance. *Management Science*, *61*(1), 146–160.

Adhvaryu, A., Kala, N., & Nyshadham, A. (2022). Management and Shocks to Worker Productivity. *Journal of Political Economy*, *130*(1), 1–47.

Adhvaryu, A., Nyshadham, A., & Tamayo, J. (2023). Managerial quality and productivity dynamics. *The Review of Economic Studies*, *90*(4), 1569–1607.

Adhvaryu, A., Nyshadham, A., & Tamayo, J. (2024). An anatomy of performance monitoring. *Harvard Business School Working Paper No. 22-066*.

Alfitian, J., Sliwka, D., & Vogelsang, T. (2024). When bonuses backfire: Evidence from the workplace. *Management Science*, *70*(9), 6395–6414.

Alfitian, J., & Vogelsang, T. (2026). The hidden costs of dismissal: Behavioral consequences of impending termination. *Journal of Labor Economics*, Forthcoming.

Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., & Vitali, A. (2020). Tackling youth unemployment: Evidence from a labor market experiment in Uganda. *Econometrica*, *88*(6), 2369–2414.

Alfonsi, L., Bassi, V., Rasul, I., & Spadini, E. (2024). The returns to skills during the pandemic: Experimental evidence from Uganda. *NBER Working Paper Series No. 32785*.

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, *104*(10), 3003–3037.

Altamuro, J., & Beatty, A. (2010). How does internal control regulation affect financial reporting? *Journal of Accounting and Economics*, *49*(1-2), 58–74.

Ambuehl, S., Niederle, M., & Roth, A. E. (2015). More money, more problems? Can high pay be coercive and repugnant? *American Economic Review*, *105*(5), 357–360.

Arellano-Bover, J. (2022). The effect of labor market conditions at entry on workers' long-term skills. *Review of Economics and Statistics*, *104*(5), 1028–1045.

Arnold, M. C., Shi, B., Tafkov, I., & Voermans, E. (2023). The effects of relative performance information and training type on employees' skill development: An experimental investigation. *Available at SSRN 4320039*.

Arnold, M. C., Shi, B., Tafkov, I., & Voermans, E. (2025). Managers' usage of management control instruments and employees' engagement in skill development. *Available at SSRN 5127947*.

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, *29*(3), 155–173.

Ashbaugh-Skaife, H., Collins, D. W., Kinney Jr, W. R., & LaFond, R. (2009). The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research*, *47*(1), 1–43.

Autor, D., Chin, C., Salomons, A., & Seegmiller, B. (2024). New frontiers: The origins and content of new work, 1940–2018. *The Quarterly Journal of Economics*, *139*(3), 1399–1465.

Autor, D., Levy, F., & Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333.

Bachmann, R., Demir, G., Green, C. P., & Uhlendorff, A. (2022). The role of within-occupation task changes in wage development. *Ruhr Economic Papers, No. 975, RWI Leibniz-Institut Für Wirtschaftsforschung*.

Bandiera, O., Barankay, I., & Rasul, I. (2011). Field experiments with firms. *Journal of Economic Perspectives*, *25*(3), 63–82.

Banker, R. D., Chang, H., & Pizzini, M. J. (2004). The balanced scorecard: Judgmental effects of performance measures linked to strategy. *The Accounting Review*, *79*(1), 1–23.

Bapna, R., Langer, N., Mehra, A., Gopal, R., & Gupta, A. (2013). Human capital investments and employee performance: An analysis of IT services industry. *Management Science*, *59*(3), 641–658.

Bartel, A. P. (1989). Formal employee training programs and their impact on labor productivity: Evidence from a human resources survey. *NBER Working Paper Series No. 3026*.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference* *to education*. University of Chicago Press.

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, *75*(299), 493–517.

Becker, G. S., & Murphy, K. M. (1988). A theory of rational addiction. *Journal of Political Economy*, *96*(4), 675–700.

Bellora-Bienengräber, L., Dalton, D. W., & McPhee, G. P. (2023). Enabling budgeting and employee turnover intentions: The role of error management climate and error aversion climate. *Available at SSRN 4358459*.

Bénabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, *70*(3), 489–520.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*(5), 1652–1678.

Berge, L. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: The R package FENmlm. *DEM Discussion Paper Series No. 13*.

Beshears, J., Lee, H. N., Milkman, K. L., Mislavsky, R., & Wisdom, J. (2021). Creating exercise habits using incentives: The trade-off between flexibility and routinization. *Management Science*, *67*(7), 4139–4171.

BIBB. (2023). *BIBB–BAuA QuBe–basisprojektion dossier: Informatik-, informations- und kommunikationstechnologieberufe*. Federal Institute for Vocational Education; Training.

Black, S. E., & Lynch, L. M. (1996). Human-capital investments and productivity. *The American Economic Review*, *86*(2), 263–267.

Bloom, N., Sadun, R., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, *122*(4), 1351–1408.

Bloom, N., Sadun, R., & Van Reenen, J. (2010). Recent advances in the empirics of organizational economics. *Annual Review of Economics*, *2*(1), 105–137.

Bloom, N., Sadun, R., & Van Reenen, J. (2012). The organization of firms across countries. *The Quarterly Journal of Economics*, *127*(4), 1663–1705.

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *Review of Economic Studies*, *91*(6), 3253–3285.

Brink, A. G., Lowe, D. J., & Victoravich, L. M. (2013). The effect of evidence strength and internal rewards on intentions to report fraud in the dodd-frank regulatory environment. *Auditing: A Journal of Practice & Theory*, *32*(3), 87–104.

Brüggen, A. (2011). Ability, career concerns, and financial incentives in a multi-task setting. *Journal of Management Accounting Research*, *23*(1), 211–229.

Brunello, G. (2009). The effect of economic downturns on apprenticeships and initial workplace training: A review of the evidence. *Empirical Research in Vocational Education and Training*, *1*(2), 145–171.

Brunello, G., & Wruuck, P. (2021). Skill shortages and skill mismatch: A review of the literature. *Journal of Economic Surveys*, *35*(4), 1145–1167.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, *108*, 43–47.

Buell, R., Cai, W., & Sandino, T. (2022). Learning or playing? The effect of gamified training on performance. *Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 19-101*.

Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2013). Overconfidence and Social Signalling. *The Review of Economic Studies*, *80*(3), 949–983.

Butts, K., & Gardner, J. (2021). {did2s}: Two-stage difference-in-differences. *arXiv Preprint arXiv:2109.05913*.

Buyalskaya, A., Ho, H., Milkman, K. L., Li, X., Duckworth, A. L., & Camerer, C. (2023). What can machine learning teach us about habit formation? Evidence from exercise and hygiene. *Proceedings of the National Academy of Sciences*, *120*(17), 1–7.

Cai, W., Chen, Y., & Chen, Y. (2024). Empowered learning as a control system. *Working Paper*.

Cai, W., Mahlendorf, M. D., & Wu, F. (2024). Supervisor behavior in times of external performance pressure. *Available at SSRN 4827928*.

Caicedo, S., Espinosa, M., & Seibold, A. (2022). Unwilling to train?—firm responses to the colombian apprenticeship regulation. *Econometrica*, *90*(2), 507–550.

Campbell, D., Epstein, M. J., & Martinez-Jerez, F. A. (2011). The learning effects of monitoring. *The Accounting Review*, *86*(6), 1909–1934.

Campos-Mercade, P., Meier, A. N., Schneider, F. H., & Weber, R. A. (2025). *What money shouldn't buy? Measuring aversion to monetary incentives for health behaviors*. Department of Economics-University of Zurich.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, *69*(5), 1127–1160.

Cardinaels, E., & Yin, H. (2015). Think twice before going for incentives: Social norms and the principal's decision on compensation contracts. *Journal of Accounting Research*, *53*(5), 985–1015.

Carrera, M., Royer, H., Stehr, M., & Sydnor, J. (2020). The structure of health incentives: Evidence from a field experiment. *Management Science*, *66*(5), 1890–1908.

Casas-Arce, P., Deller, C., Martínez-Jerez, F. A., & Narciso, J. M. (2023). Knowing that you know: Incentive effects of relative performance disclosure. *Review of Accounting Studies*, *28*(1), 91–125.

Casas-Arce, P., Martínez-Jeres, F. A., & Moran, J. (2024). Motivating from the heights: A field experiment on top managers visiting the front-line. *Review of Accounting Studies*, *30*(23), 1099–1135.

Casas-Arce, P., & Martinez-Jerez, F. A. (2009). Relative performance compensation, contests, and dynamic incentives. *Management Science*, *55*(8), 1306–1320.

Casas-Arce, P., Martínez-Jerez, F. A., & Moran, J. (2025). Motivating from the heights: A field experiment on top managers visiting the front-line: P. Casas-arce et al. *Review of Accounting Studies*, *30*(2), 1099–1135.

Castro, S., Ho, Hoa, & Mickeler, M. (2025). Making help visible: Experimental evidence from a recognition program in the workplace. *Available at SSRN 5223745*.

Cederlöf, J., Fredriksson, P., Nekoei, A., & Seim, D. (2024). Mandatory notice of layoff, job search, and efficiency. *The Quarterly Journal of Economics*, *140*(1), 585–633.

Celhay, P. A., Gertler, P. J., Giovagnoli, P., & Vermeersch, C. (2019). Long-run effects of temporary incentives on medical care productivity. *American Economic Journal: Applied Economics*, *11*(3), 92–127.

Chalmers, K., Hay, D., & Khlif, H. (2019). Internal control in accounting research: A review. *Journal of Accounting Literature*, *42*(1), 80–103.

Chapman, G., & Gneezy, U. (2024). Change behavior, motivation will follow (?): Acquired taste and incentives. *AEA Papers and Proceedings*, *114*, 660–665.

Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica*, *77*(3), 909–931.

Choi, J., Hecht, G. W., & Tayler, W. B. (2012). Lost in translation: The effects of incentive compensation on strategy surrogation. *The Accounting Review*, *87*(4), 1135–1163.

Choi, J., Hecht, G. W., & Tayler, W. B. (2013). Strategy selection, surrogation, and strategic performance measurement systems. *Journal of Accounting Research*, *51*(1), 105–133.

Christ, M. H. (2013). An experimental investigation of the interactions among intentions, reciprocity, and control. *Journal of Management Accounting Research*, *25*(1), 169–197.

Cronin, M., Erkens, D. H., Schloetzer, J. D., & Tinsley, C. H. (2021). How controlling failure perceptions affects performance: Evidence from a field experiment. *The Accounting Review*, *96*(2), 205–230.

Danilov, A., & Sliwka, D. (2017). Can contracts signal social norms? Experimental evidence. *Management Science*, *63*(2), 459–476.

De Charms, R. (1968). *Personal causation: The internal affective determinants of behavior*.

Routledge.

De Grip, A., & Sauermann, J. (2012). The effects of training on own and co-worker productivity: Evidence from a field experiment. *The Economic Journal*, *122*(560), 376–399.

De Janvry, A., He, G., Sadoulet, E., Wang, S., & Zhang, Q. (2023). Subjective Performance Evaluation, Influence Activities, and Bureaucratic Work Behavior: Evidence from China. *American Economic Review*, *113*(3), 766–799.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105.

Deci, E. L., Connell, J. P., & Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology*, *74*(4), 580.

Delfino, A., Garnero, A., Inferrera, Sergio, Leonardi, M., & Sadun, R. (2026). Unwilling to reskill? Experimental evidence from real-word jobseekers. *NBER Working Paper No. 34633*.

Deming, D. J. (2022). Four facts about human capital. *Journal of Economic Perspectives*, *36*(3), 75–102.

DeStatis. (2025). *Daten zu den indikatoren des deutschlandatlas interaktiv*. Statistisches Bundesamt. https://www.deutschlandatlas.bund.de/SharedDocs/Downloads/DE/Deutschlandatlas-Daten.html

Diaz, B. S., Nazarrett, A. N., Ramirez, J., Sadun, R., & Tamayo, J. A. (2025). Training within firms. *NBER Working Paper Series No. 33670*.

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *308*(1135), 67–78.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, *14*(1), 57–74.

Espinosa, M., & Stanton, C. T. (2023). Training, communications patterns, and spillovers inside organizations. *NBER Working Paper Series No. 30224*.

Ewers, M., & Zimmermann, F. (2015). Image and Misreporting. *Journal of the European Economic Association*, *13*(2), 363–380.

Eyring, H. (2020). Disclosing physician ratings: Performance effects and the difficulty of altering ratings consensus. *Journal of Accounting Research*, *58*(4), 1023–1067.

Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, *96*(5), 1611–1630.

Feichter, C., & Grabner, I. (2020). Empirische forschung zu management control –

ein überblick und neue trends. *Schmalenbachs Zeitschrift Für Betriebswirtschaftliche Forschung*, *72*(2), 149–181.

Feltham, G. A., & Xie, J. (1994). Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review*, *69*(3), 429–453.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *The Journal of Abnormal and Social Psychology*, *58*(2), 203.

Field, E. M., Linden, L. L., Malamud, O., Rubenson, D., & Wang, S.-Y. (2019). *Does vocational education work? Evidence from a randomized experiment in mongolia*. NBER Working Paper No. 26092.

Fischer, S., Frese, M., Mertins, J. C., & Hardt-Gawron, J. V. (2018). The role of error management culture for firm and individual innovativeness. *Applied Psychology*, *67*(3), 428–453.

Frese, M., & Keith, N. (2015). Action errors, error management, and learning in organizations. *Annual Review of Psychology*, *66*(1), 661–687.

Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American Economic Review*, *87*(4), 746–755.

Friebel, G., Heinz, M., Krueger, M., & Zubanov, N. (2017). Team incentives and performance: Evidence from a retail chain. *American Economic Review*, *107*(8), 2168–2203.

Friebel, G., Heinz, M., & Zubanov, N. (2022). Middle managers, personnel turnover, and performance: A long-term field experiment in a retail chain. *Management Science*, *68*(1), 211–229.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836.

Gallani, S. (2023). Conduit incentives: Eliciting cooperation from workers outside of managers' control. *The Accounting Review*, *98*(3), 175–202.

Gardner, B., Abraham, C., Lally, P., & Bruijn, G.-J. de. (2012). Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the self-report habit index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*, 1–12.

Gardner, B., Bruijn, G.-J. de, & Lally, P. (2011). A systematic review and meta-analysis of applications of the self-report habit index to nutrition and physical activity behaviours. *Annals of Behavioral Medicine*, *42*(2), 174–187.

Gardner, B., Rebar, A. L., & Lally, P. (2022). How does habit form? Guidelines for tracking real-world habit formation. *Cogent Psychology*, *9*(1), 2041277.

Gardner, J. (2022). Two-stage differences in differences. *arXiv Preprint arXiv:2207.05943*.

Gertler, P., Higgins, S., Scott, A., & Seira, E. (2018). *The long-term effects of temporary incentives to save: Evidence from a prize-linked savings field experiment*. Abdul Latif Jameel Poverty Action Lab.

Gibbs, M., & Bazylik, S. (2022). *How is new technology changing job design?* (p. 344). IZA World of Labor, Institute of Labor Economics (IZA).

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(3), 523–536.

Giuntella, O., Saccardo, S., & Sadoff, S. (2024). Sleep: Educational impact and habit formation. *NBER Working Paper No. 32550*.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, *25*(4), 191–210.

Gneezy, U., & Rustichini, A. (2000a). A fine is a price. *The Journal of Legal Studies*, *29*(1), 1–17.

Gneezy, U., & Rustichini, A. (2000b). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, *115*(3), 791–810.

Gosnell, G. K., List, J. A., & Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy*, *128*(4), 1195–1233.

Grabe, L. (2025). Employee performance in response to workplace errors: Evidence from the field. *Unpublished Working Paper*.

Grabe, L., & Sliwka, D. (2025). Managing skills in organizations - evidence from a field experiment. *IZA Discussion Paper No 17727*.

Gravert, C., & Collentine, L. O. (2021). When nudges aren't enough: Norms, incentives and habit formation in public transport usage. *Journal of Economic Behavior & Organization*, *190*, 1–14.

Grundke, R., Marcolin, L., Nguyen, The Linh Bao, & Squicciarini, M. (2018). Getting skills right: Skills for jobs indicators. *OECD Science, Technology and Industry Working Papers, No. 2018/09*.

Guo, J., Huang, P., Zhang, Y., & Zhou, N. (2016). The effect of employee treatment policies on internal control weaknesses and financial restatements. *The Accounting Review*, *91*(4), 1167–1194.

Hamilton, E. L., & Smith, J. L. (2021). Error or fraud? The effect of omissions on management's fraud strategies and auditors' evaluations of identified misstatements. *The*

Accounting Review, *96*(1), 225–249.

Hannan, R. L., Krishnan, R., & Newman, A. H. (2008). The effects of disseminating relative performance feedback in tournament and individual performance compensation plans. *The Accounting Review*, *83*(4), 893–913.

Hannan, R. L., McPhee, G. P., Newman, A. H., & Tafkov, I. D. (2013). The effect of relative performance information on performance and effort allocation in a multi-task environment. *The Accounting Review*, *88*(2), 553–575.

Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, *73*, 103–130.

Harmon-Jones, E., & Mills, J. (2019). *An introduction to cognitive dissonance theory and an overview of current perspectives on the theory.* American Psychological Association.

Heese, J., & Pérez-Cavazos, G. (2020). When the boss comes to town: The effects of headquarters' visits on facility-level misconduct. *The Accounting Review*, *95*(6), 235–261.

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1-3), 337–386.

Hoffman, M., & Stanton, C. T. (2024). People, practices, and productivity: A review of new advances in personnel economics. *NBER Working Paper Series No. 32849*.

Holmstrom, B. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*(Special Issue), 24–52.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, *10*(1), 74.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, *66*(1), 169–182.

Holmström, B. (2017). Pay for performance and beyond. *American Economic Review*, *107*(7), 1753–1777.

Huffman, D., & Bognanno, M. (2018). High-powered performance pay and crowding out of nonmonetary motives. *Management Science*, *64*(10), 4669–4680.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory.* Appleton-Century.

Jackson, C. K., & Schneider, H. S. (2015). Checklists and Worker Behavior: A Field Experiment. *American Economic Journal: Applied Economics*, *7*(4), 136–168.

Jensen, N., Lyons, E., Chebelyon, E., Bras, R. L., & Gomes, C. (2020). Conspicuous monitoring and remote work. *Journal of Economic Behavior & Organization*, *176*, 489–511.

Jovanovic, B., & Nyarko, Y. (1996). Learning by doing and the choice of technology.

Econometrica, *64*(6), 1299–1310.

Kantor, J., & Sundaram, A. (2022). *The rise of the worker productivity score. The New York Times*. https://www.nytimes.com/interactive/2022/08/14/business/worker-productivity-tracking.html

Katok, E., & Siemsen, E. (2011a). The influence of career concerns on task choice: Experimental evidence. *Management Science*, *57*(6), 1042–1054.

Katok, E., & Siemsen, E. (2011b). Why genius leads to adversity: Experimental evidence on the reputational effects of task difficulty choices. *Management Science*, *57*(6), 1042–1054.

Kowaleski, Z. T., Sutherland, A. G., & Vetter, F. W. (2024). The effect of supervisors on employee misconduct. *The Accounting Review*, *99*(3), 287–313.

Krueger, M., & Friebel, G. (2022). A pay change and its long-term consequences. *Journal of Labor Economics*, *40*(3), 543–572.

Larcom, S., Rauch, F., & Willems, T. (2017). The benefits of forced experimentation: Striking evidence from the london underground network. *The Quarterly Journal of Economics*, *132*(4), 2019–2055.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, *90*(5), 1346–1361.

Lazear, E. P. (2018). Compensation and incentives in the workplace. *Journal of Economic Perspectives*, *32*(3), 195–214.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the" overjustification" hypothesis. *Journal of Personality and Social Psychology*, *28*(1), 129.

Lipowski, C. (2024). No teens, no tech: How shortages of young workers hinder firm technology investments. *ZEW Discussion Papers*.

Liu, L., Wang, Y., & Xu, Y. (2024). A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *American Journal of Political Science*, *68*(1), 160–176.

Loewenstein, G., & Angner, E. (2003). Predicting and indulging changing preferences. In G. Loewenstein, D. Read, & R. F. Baumeister (Eds.), *Time and decision: Economic and psychological perspectives on intertemporal choice* (pp. 351–391). Russell Sage Foundation New York.

Lu, J. (2022). Limited attention: Implications for financial reporting. *Journal of Accounting Research*, *60*(5), 1991–2027.

Mahlendorf, M. D., & Vogelsang, T. (2024). Performance measures and incentives: evidence from field experiments in firms. In A. M. Lillis & J. Grafton (Eds.), *Research Handbook on Performance Measurement for Management Control* (pp. 155–166). Edward Elgar Publishing.

Majumdar, S. (2007). Market conditions and worker training: How does it affect and whom? *Labour Economics*, *14*(1), 1–23.

Maletta, M., & Wright, A. (1996). Audit evidence planning: An examination of industry error characteristics. *Auditing: A Journal of Theory & Practice*, *15*(1), 71.

Manthei, K., & Sliwka, D. (2019). Multitasking and subjective performance evaluations: Theory and evidence from a field experiment in a bank. *Management Science*, *65*(12), 5861–5883.

Manthei, K., Sliwka, D., & Vogelsang, T. (2021). Performance pay and prior learning—evidence from a retail chain. *Management Science*, *67*(11), 6998–7022.

Manthei, K., Sliwka, D., & Vogelsang, T. (2023). Talking about performance or paying for it? A field experiment on performance reviews and incentives. *Management Science*, *69*(4), 2198–2216.

Marshall, A. (1920). *Principles of economics* (8th ed.). Macmillan.

Mazar, A., & Wood, W. (2018). Defining habit in psychology. In B. Verplanken (Ed.), *The psychology of habit: Theory, mechanisms, change, and contexts* (pp. 13–29). Springer.

Melchionne, K. (2007). Acquired taste. *Contemporary Aesthetics (Journal Archive)*, *5*(Article 11).

Merchant, K., & Van der Stede, W. (2017). *Management control systems - performance measurement, evaluation and incentives*. Pearson.

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*(1), 465–489.

Milgrom, P., & Roberts, J. (1992). *Economics, organization and management* (Vol. 7). Prentice-hall Englewood Cliffs, NJ.

Mincer, J. (1974). The human capital earnings function. In *Schooling, experience, and earnings* (pp. 83–96). NBER.

Moers, F. (2024). MAS & IMA doctoral colloquium: Management accounting and labor economics. *American Accounting Association Management Accouting Section*.

Murty, R. N., & Karanth, S. (2022). *Monitoring individual employees isn't the way to boost productivity. Harvard Business Review*. https://hbr.org/2022/10/monitoring-individual-employees-isnt-the-way-to-boost-productivity

Nebeker, D. M., & Tatum, B. C. (1993). The effects of computer monitoring, standards, and rewards on work performance, job satisfaction, and stress. *Journal of Applied Social Psychology*, *23*(7), 508–536.

Owhoso, V. E., Messier, J. W. F., & Lynch, J. J. G. (2002). Error detection by industry-specialized teams during sequential audit review. *Journal of Accounting Research*, *40*(3), 883–900.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, *37*(1), 7–63.

Ramsay, R. J. (1994). Senior/manager differences in audit workpaper review performance. *Journal of Accounting Research*, *32*(1), 127–135.

Ravid, D. M., White, J. C., Tomczak, D. L., Miles, A. F., & Behrend, T. S. (2023). A meta-analysis of the effects of electronic performance monitoring on work outcomes. *Personnel Psychology*, *76*(1), 5–40.

Rebbitt, D. (2014). Pyramid power: A new view of the great safety pyramid. *Professional Safety*, *59*(09), 30–34.

Reljić, J., Evangelista, R., & Pianta, M. (2021). Digital technologies, skill upgrading and employment: Evidence from european firms. *Industrial and Corporate Change*, *30*(3), 643–668.

Rochambeau, G. de. (2017). Monitoring and intrinsic motivation: Evidence from Liberia's trucking firms. *International Growth Center Working Paper No. F-51303-LIB-1*.

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *The American Economic Review*, *63*(2), 134–139.

Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, *235*(2), 2218–2244.

Royer, H., Stehr, M., & Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics*, *7*(3), 51–84.

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, *45*(4), 736–750.

Sandvik, J., Saouma, R. E., Seegert, N. T., & Stanton, C. T. (2020). Workplace knowledge flows. *The Quarterly Journal of Economics*, *135*(3), 1635–1680.

Sandvik, J., Saouma, R. E., Seegert, N., & Stanton, C. (2025). Should human capital development programs be mandatory or voluntary? Evidence from a field experiment on mentorship. *Management Science*, Forthcoming.

Sandvik, J., Saouma, R., Seegert, N., & Stanton, C. (2021). Employee responses to compensation changes: Evidence from a sales firm. *Management Science*, *67*(12), 7687–7707.

Schmidheiny, K., & Siegloch, S. (2023). On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization. *Journal of Applied Econometrics*, *38*(5), 695–713.

Shrikant, A. (2023). *Companies use AI to monitor workers and 45% of employees say it has a negative effect on their mental health. CNBC Make IT*. https://www.cnbc.com/2023/09/08/employers-using-ai-to-monitor-workers-has-negative-impact-on-employees.html

Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, *88*(4), 641–658.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Sliwka, D. (2007). Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review*, *97*(3), 999–1012.

Stigler, G. J., & Becker, G. S. (1977). De gustibus non est disputandum. *American Economic Review*, *67*(2), 76–90.

Stüber, R. (2024). Why high incentives cause repugnance: A framed field experiment. *The Economic Journal*, *134*(662), 2580–2620.

Van Dyck, C., Frese, M., Baer, M., & Sonnentag, S. (2005). Organizational error management culture and its impact on performance: A two-study replication. *Journal of Applied Psychology*, *90*(6), 1228.

Verplanken, B., & Orbell, S. (2003). Reflections on past behavior: A self-report index of habit strength. *Journal of Applied Social Psychology*, *33*(6), 1313–1330.

Verplanken, B., & Orbell, S. (2022). Attitudes, habits, and behavior change. *Annual Review of Psychology*, *73*(1), 327–352.

Volpp, K. G., & Loewenstein, G. (2020). What is a habit? Diverse mechanisms that can produce sustained behavior change. *Organizational Behavior and Human Decision Processes*, *161*, 36–38.

Wang, K. K., Cheng, M. M., & Chang, L. J. (2023). Reducing strategy surrogation: The effects of performance measurement system flexibility and environmental dynamism. *The Accounting Review*, *98*(4), 435–456.

Wood, W., & Neal, D. T. (2009). The habitual consumer. *Journal of Consumer Psychology*, *19*(4), 579–592.

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, *83*(6), 1281.

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*, 289–314.

Wu, H. X., Liu, S. X., & Xu, H. (2024). Managerial attention, employee attrition, and productivity: Evidence from a field experiment. *Rotman School of Management Working Paper No. 3787204*.

Yang, N., & Long Lim, Y. (2018). Temporary incentives change daily routines: Evidence from a field experiment on singapore's subways. *Management Science*, *64*(7), 3365–3379.

Eidesstattliche Versicherung

gemäß § 26 Abs. 5 der Promotionsordnung im Promotionsprogramm Research in Management, Economics and Social Sciences der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln vom 11. August 2025:

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Weitere Personen, neben den ggf. in der Arbeit aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten fremden Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von den angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist, sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen der Promotionsordnung im Promotionsprogramm Research in Management, Economics and Social Sciences derWirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Leitlinien der Universität zu Köln zur Sicherung guter wissenschaftlicher Praxis gelesen und sie bei der Durchführung der der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Leonhard Grabe
Köln, 23.01.2026