# The Polarizing Performance Effect of Private Social Comparison Information

Maximilian Kohler♣ Matthias D. Mahlendorf♠

Mischa Seiter♥ Timo Vogelsang♦

*This Version: February 4th, 2024*

**Abstract**

We conduct a field experiment in a retail chain to study the effects of two common private relative performance information (RPI) designs on employees' performance. Store employees are randomly assigned to one of three groups: a control group receiving no RPI, a treatment group receiving RPI based on other employees' aggregate performance (median), or a treatment group receiving the RPI based on performance distribution (deciles). Our setting does not provide financial incentives. On average, we do not find statistically significant treatment effects on performance. However, in line with economic theory on social comparison and as hypothesized, we find that private RPI can lead to polarization. While employees at the top of the performance distribution increase their performance after receiving private RPI, the effect on low and medium performers is non-significantly negative. This polarization effect is only present for distribution RPI and not for aggregate RPI.

Key words: Social Comparison, Relative Performance Information, Performance Distribution, Field Experiment

JEL: M4, M40, M46, M1, M12

Acknowledgments: We thank Isabella Grabner, Konstantin Gruen, Sabra Khajehnejad, Laurence van Lent, and Gerhard Speckbacher for their helpful comments. Furthermore, we thank the participants at the research seminars at Vrije Universiteit Amsterdam, KU Leuven, and Frankfurt School of Finance and Management as well as conference participants at ERMAC and EIASM for their insightful feedback. We are very grateful for the assistance from the company. The workers' council of the retail chain and the ethics committee of Frankfurt School of Finance & Management approved the experiment. The experiment was pre-registered with the ID AEARCTR – 0008463.

Competing Interests Statement: Maximilian Kohler is working at the company that is subject of the study. In his role as an authorized signatory, he is responsible for management accounting, human resources, and marketing. However, no funding was received from the company, and none of the results were adjusted to the company's demand. Any errors and all opinions are our own.

♣ Maximilian Kohler, Ulm University, Institute of Business Analytics, Helmholtzstraße 22 Ulm, Germany, E-Mail: maximilian.kohler@uni-ulm.de

♠ Matthias D. Mahlendorf, Frankfurt School of Finance & Management, Department of Accounting, Adickesallee 32-34, Frankfurt Germany, E-Mail: m.mahlendorf@fs.de, tel. +49 69 154008 – 837

♥ Mischa Seiter, Ulm University, Institute of Business Analytics, Helmholtzstraße 22 Ulm, Germany, E-Mail: mischa.seiter@uni-ulm.de, +49 73150 – 32307

♦ Timo Vogelsang, Frankfurt School of Finance & Management, Department of Accounting, Adickesallee 32-34, Frankfurt, Germany, E-Mail: t.vogelsang@fs.de, tel. +49 69 154008 – 393

# 1. Introduction

Laboratory experiments have shown that enabling people to anonymously compare themselves with others can affect their performance in the absence of incentives (e.g., Tafkov, 2013). However, little evidence exists whether private relative performance information (RPI) affects the performance of employees in the workplace. Moreover, the few existing field experiments that cooperate with firms mostly introduce private RPI in settings where performance-based pay is present (e.g., Ashraf, 2022; Barankay, 2012). In such settings, learning about one's relative position also provides information about the possibility to earn a higher pay, even if pay for performance is based on absolute and not relative performance. Building on economic theory on utility derived from social comparison, we study the effects of two common RPI designs, which are expected to have important implications for the performance effect of RPI – aggregate RPI and RPI on the performance distribution. To the best of our knowledge, our field experiment is among the first to implement private RPI on performance against an untreated control group within a firm that offers no incentives for performance improvements. This enables us to study the isolated effect of social comparison on employee performance.[1]

Our sample includes 596 sales employees of a German supermarket chain.[2] Specifically, we provide employees at fresh food counters with private RPI on their average sales per transaction using a digital app on a weekly basis. There is a no RPI condition (*Control*) and two treatment groups (*Median RPI* and *Distribution RPI*). Using clustered stratified randomization, we assign employees to one of the three groups at the store level. Employees in the *Control* group receive information about their average sales per transaction.[3] Employees in the *Median RPI* group also receive information about their average sales per transaction and, in addition, the median average sales per transaction of employees from similar departments of comparable stores. Employees in the *Distribution RPI* group receive information about their average sales per transaction and, in addition, the average sales per transaction of the deciles ($10^{th}$, $20^{th}$, ..., $100^{th}$) from similar departments of comparable stores.

[1] This is an important distinction due to the positive effect of performance-based pay on performance (e.g., Banker et al., 1996; Lazear, 2000; Lourenço, 2016; Manthei et al., 2023). Firm level field experiments combine the advantages of control and realism and are a small but growing literature (e.g., Belnap, 2023; Blanes i Vidal & Nossol, 2011; Costello et al., 2020; Floyd & List, 2016).

[2] The experiment was pre-registered with the ID AEARCTR – 0008463.

[3] The control group receives this information to avoid contamination due to the information provided and to avoid that our treatment effects are driven by attention on a specific performance metric (Manthei et al., 2023).

In general, social comparison theory suggests that private RPI motivates employees to increase performance as people strive to outperform others (Suls & Wheeler, 2000). This is supported by evidence from the lab, showing that private RPI positively affects performance (e.g., Tafkov, 2013). However, mixed results from recent field experiments testing private RPI in company settings question whether this holds in practice. While some studies find that private RPI positively affects performance (e.g., Blanes i Vidal & Nossol, 2011), others suggest that it does not have an effect (Ashraf, 2022) or that private RPI risks demoralizing employees (Barankay, 2012). Importantly, no study yet tests the performance effect of providing employees with private RPI on their performance in the absence of incentives in a company setting. Consequently, our analysis starts with private RPI's average performance effect in the absence of incentives.

This is followed by an analysis of pre-registered heterogeneous treatment effects of private RPI.[4] While private RPI might motivate employees who expect to be able to outperform their peers, it could demotivate employees who are far below the reference point (Eyring & Narayanan, 2018; Roels & Su, 2014). The resulting polarization effects could be especially severe when RPI is presented as an aggregate performance measure such as *Median RPI*. With an aggregate reference point, the performance distance between an employee's performance and the reference point can be rather wide, risking to demotivate low performers (Roels & Su, 2014). With RPI in the form of the reference distribution (*Distribution RPI*), on the other hand, employees have multiple reachable reference points close to them, which could reduce potential polarization effects. This is important because higher performance variance might allow identifying top performers but, at the same time, risks demoralizing low performers (Roels & Su, 2014).

Our field experiment provides three key results. First, in contrast to our pre-registration, we do not find that providing private RPI, on average, affects performance compared to an untreated control group. Second, in line with our pre-registration, we find that the missing average effect is partially explained by heterogeneous effects that offset each other. Our results show that the effect of *Distribution RPI* is significantly positive for employees at the top of the performance distribution but non-significantly negative for low and medium performers, thereby causing a (one-sided) polarization effect. Third, in contrast to our pre-registration, the polarization effect is not more severe for aggregate RPI (*Distribution RPI*) than for RPI in the

[4] Precisely, we stated in our pre-registration: "[…] we expect that providing information on peer performance will increase performance variance.".

form of the reference distribution (*Median RPI)*. If anything, the opposite is the case, as the effects are only statistically significant for *Distribution RPI*, and not for *Median RPI*.

Our study makes two main contributions. (1) To the best of our knowledge, we are the first to implement a field experiment in a company setting where private social comparison information on performance is studied against a no RPI group in the absence of incentives. The results indicate that such an isolated social comparison does not increase performance on average. However, it does for top performers.

(2) Our field experiment contributes to current research testing mechanisms to avoid potential negative effects of RPI (Hermes et al., 2021; Rilke et al., 2021). It is the first field experiment that contrasts the two common RPI designs *Median* RPI versus *Distribution RPI*, which has been predicted to be an important choice in economic theory on social comparison (Roels & Su, 2014). The importance is further highlighted by various studies that use either aggregate RPI (e.g., Azmat & Iriberri, 2010; Brade et al., 2023; Eyring & Narayanan, 2018) or detailed RPI on the performance distribution (e.g., Barankay, 2012; Blanes i Vidal & Nossol, 2011; Kohler et al., 2023). In contrast to the propositions of Roels & Su' (2014), we find that the polarization effect of RPI is not avoided by providing RPI in the form of the reference distribution. If anything, the opposite is the case. A practical implication is that if social designers are interested in increasing the heterogeneity of the reference group (Roels & Su, 2014), for example to identify top performers, they should rather implement *Distribution* RPI than *Median* RPI in settings like ours.

## 2. Literature Overview and Hypothesis

Generally, social comparison theory assumes that, in the absence of objective standards, people tend to compare themselves to others to evaluate their performance (Festinger, 1954; Suls & Wheeler, 2000). In this regard, social comparison theory suggests that outperforming others provides individuals with positive utility as it enhances their self-image (Smith, 2000). Relative performance information (RPI) informs individuals about their performance compared to others. Thus, when presented with RPI, employees are expected to increase performance even without performance-based incentives (Tafkov, 2013).

This study focuses on RPI that is only provided to the individuals but not made public or accessed by supervisors (private RPI). Although the performance effect of RPI is stronger when the names of individuals are public (e.g., Tafkov, 2013), private RPI might be easier to implement in the field concerning to data protection laws and potential resistance from employee representatives, making it a relevant subject for research. Contrasting evidence from the lab, which typically finds a motivating effect for private RPI (e.g., Hannan et al., 2013; Tafkov, 2013), field research in education settings and companies provides mixed results.[5] Some studies find that private RPI positively affects performance (Azmat & Iriberri, 2010; Blanes i Vidal & Nossol, 2011; Eyring & Narayanan, 2018; Tran & Zeckhauser, 2012). Others suggest that private RPI does not affect performance (Ashraf, 2022) or risks reducing performance because it demotivates low performers and employees that perform worse than expected (Ashraf et al., 2014; Barankay, 2012). Still others investigate specific design measures to avoid potentially negative effects of private RPI. In this regard, Rilke et al. (2021) find that RPI on effort instead of performance motivates low performers.

Furthermore, few studies implement private RPI in a company setting (Ashraf, 2022; Barankay, 2012; Blanes i Vidal & Nossol, 2011; Kohler et al., 2023; Rilke et al., 2021).[6] Except for Rilke et al. (2021), who test private RPI based on an input measure, and thus on effort instead of performance, there is no field experiment yet that tests the effect of private RPI compared to a no RPI condition in the absence of confounding factors such as financial incentives. For example, Barankay (2012) and Ashraf (2022) both provide piece rate incentives which potentially distort the effect of social comparison (Tafkov, 2013). On the one hand, under financial incentives, RPI might motivate employees because it informs them about potential payoffs, not due to social comparison motives. On the other hand, employees may already provide their maximum effort, reducing the opportunity for further performance improvement through RPI.

Considering the potential biases due to confounding factors in existing field studies, further research is needed on the isolated effect of private RPI on performance. Being able to test private RPI in a company setting that does not provide explicit or implicit incentives, we thus start with our basic hypothesis, which is in line with social comparison theory:

[5] Given the extensive literature on performance feedback (e.g., Casas-Arce et al., 2017) and social comparison theory (e.g., Suls & Wheeler, 2000), we focus our literature review on studies in accounting and economics journals that have tested the performance effect of RPI in the field. Furthermore, we disentangle our literature review from the tournament literature (e.g., Delfgaauw et al., 2013; Kelly et al., 2017) by focusing on studies that test the effect of RPI in the absence of explicit monetary incentives. However, we also acknowledge the extensive research done in laboratory experiments which typically find a positive performance effect (e.g., Hannan et al., 2008; Hannan et al., 2013; Kramer et al., 2016; Kuhnen & Tymula, 2012; Tafkov, 2013). For studies testing public RPI in a company setting, see Ashraf (2022); Blader et al. (2020); Eyring (2020); Song et al. (2018).

[6] Kohler et al. (2023) test the performance effect of providing RPI on separate tasks or RPI on separate tasks in combination with RPI on overall performance instead of providing only RPI on overall performance. Moreover, in this study it is not possible to evaluate the overall effect of RPI as not control group without RPI is implemented.

*H1: Introducing private RPI increases performance on average.*

Recent evidence suggests that one should consider heterogeneity when analyzing the effect of private RPI as its motivational effect differs depending on the reference point as well as one's ability to outperform it (Brade et al., 2023; Eyring & Narayanan, 2018). A reference point too high might demotivate, while a reference point too low risks losing its motivational effect (Eyring & Narayanan, 2018). In this regard, our further hypotheses build on Roels and Su (2014), who formulate a model to predict the effects of RPI depending on employees' ability to outperform the reference point. Specifically, Roels and Su (2014) predict that RPI causes polarization by motivating high performers more than low performers in settings where people gain utility by outperforming their peers but no disutility when being outperformed (ahead-seeking environment).[7]

Roels and Su (2014) assume that, in an ahead-seeking environment, the utility gained from a performance comparison depends on the positive performance distance between one's performance and the reference point. Furthermore, they assume that this relationship is linear. As high performers can expect to be able to outperform the reference point, Roels and Su (2014) predict that they increase performance to gain utility from social comparison as a result of RPI.[8] Low performers who think that they cannot outperform their peers, on the other hand, are expected to give up and reduce performance. Together, the effects on low and high performers are predicted to cause a polarization effect (see Figure 1).[9] We formulate our treatments in an ahead-seeking way (see 4.2 The Treatments). Furthermore, our setting provides the opportunity to gain utility for employees who outperform their peers but no disutility for employees who

[7] Roels and Su (2014) distinguish between ahead-seeking and behind-averse behavior. In this regard, they suggest that whether one is ahead-seeking or behind-averse can be shaped by the institutional environment. If ahead-seeking, one gains utility from overperforming relative to peers. If behind-avers, one experiences a utility loss from underperforming relative to peers. We focus on the model's predictions for heterogeneous groups in ahead-seeking environments because we consider these elements to best fit our empirical setting. We leave empirical tests of other parts of the model (e.g., behind-averse environments and homogeneous groups) for future research.

[8] In their model, Roels and Su (2014) assume that there are high performers and low performers with different lower and upper performance bounds indicated by their performance. They assume that this is not related to being ahead-seeking or behind-averse.

[9] Note that Roels and Su (2014) predict this pattern for an ahead-seeking environment. In a behind-averse environment, the game-theoretic model predicts a clustering close to the average of the outcome spectrum (Roels & Su, 2014). The intuition is that low performers will choose a high performance to avoid underperforming the reference point. In contrast, high performers will choose a low performance as this is still enough to not underperform the reference point. The propositions leading to H2 remain valid if the reference group is mixed but mainly ahead-seeking (Roels & Su, 2014). To foster an ahead-seeking environment, one can display injunctive and/or descriptive messages which focus employees' attention on outperforming others (Roels & Su, 2014; Schultz et al., 2007).

underperform their peers. Building on Roels and Su (2014), we thus test the following hypothesis:

*H2: Introducing private RPI in an ahead-seeking environment causes polarization by motivating high performers while demotivating low performers.*

While an increased performance variance allows identifying top performers, it risks demoralizing low performers, which may harm the work atmosphere and firm performance in the long run. The designer of an RPI intervention might thus want to avoid strong polarization effects. Roels and Su (2014) suggest that to that aim, RPI should be provided in the form of detailed information on the reference distribution (e.g., deciles) instead of RPI in the form of an aggregate reference point (e.g., median) (Figure 1). The intuition is that employees compare their performance to the individual reference points.[10] This allows low performers to outperform some other reference points, i.e., other low performers, motivating them to increase performance. On the other hand, high performers might be less motivated by RPI in the form of the performance distribution. While aggregate RPI provides them with constant positive feedback, RPI on the performance distribution informs them that they perform well but that there are others which they cannot outperform, reducing their motivation to further increase performance. Thus, detailed RPI on the performance distribution is expected to have a lower polarization effect than RPI as an aggregate reference point (Roels & Su, 2014). Consequently, we investigate the following hypothesis:

*H3: The polarization effect from private RPI is larger in the aggregated RPI condition than in the performance distribution condition compared to the control condition.*

*Insert Figure 1 about here.*

However, we acknowledge that H3 is not without tension. Taking into account prior research that finds a positive performance effect of private RPI among low performers (e.g., Azmat & Iriberri, 2010; Blanes i Vidal & Nossol, 2011; Rilke et al., 2021), we acknowledge that private RPI does not always demotivate low performers. As long as low performers expect that they can outperform the reference point, RPI may motivate them (Eyring & Narayanan, 2018; Roels

[10] If the distribution of reference points is shown instead of an aggregated reference point, the performance distance corresponds to the sum of the linear differences between one's performance and the outperformed reference points in an ahead-seeking environment (Roels & Su, 2014).

& Su, 2014). Additionally, considering the constraints of a company setting, low performers might be unable to further reduce their performance due to private RPI. While in theory they can give up, they must fulfill customers' or supervisors' requests at the minimum in practice. Furthermore, we acknowledge that aggregate RPI might not be more motivating to high performers than detailed RPI on the performance distribution. For example, contrasting the above argument, Eyring and Narayanan (2018) show that the positive performance effect of RPI decreases the further one outperforms the reference point, i.e., the utility gained from social comparison is not linear. Consequently, instead of demotivating high performers, detailed RPI on the performance distribution might be more motivating to them as it provides more ambitious reference points.

## 3. Field Setting

Our experiment uses data from a large German retail organization. Specifically, we focus on employee-level performance data from 596 sales employees working in 42 supermarkets located in southern Germany (one region of the company). On average, an employee generates €2,210 in sales per week in a store that is, on average, 3,086 square meters in size. In our study, we focus on each store's sales employees in the butchery, cheese, and fish departments, where customers can buy fresh goods at food counters, similar to a weekly market.

Sales employees are commonly tasked with stocking and presenting products and serving customers (i.e., selling and advising). They work in two shifts, morning and afternoon, which they switch from week to week. Employees share all the above-described tasks. They have practically no control over the number of customers who visit their department, but they do have significant impact on what customers buy. Employees can increase their average sales per transaction by selling additional products or upselling customers on higher-priced products. They can do so by giving better advice and applying sales techniques, which requires spending more time with a customer. The company aims to offer customers the best assortment as well as a high quality service. Thus, it promotes such behavior offering employees training on sales techniques and the assortment.

The company does not provide any individual financial incentives based on individual sales performance. Instead, employees receive a fixed monthly salary according to the collective wage agreement for their industry. Additionally, due to an agreement with the works council, the company is not allowed to use employee-level performance data to evaluate sales employees. Employees can advance their careers to management roles or become experts in their product category, such as a meat sommelier. However, such promotions are based on skills like personnel management, order volume estimation, and basic business knowledge rather than primarily on quantifiable sales performance. Our setting thus provides neither explicit nor implicit incentives for employees who under- or outperform their peers in sales performance.

The information provided to employees before the experiment is as follows. Employees are granted access to department-level performance measures (sales, average sales per transaction, margin) via their manager. Receipts are generated for customers after each transaction, giving employees frequent performance feedback. This information helps them understand how their effort translates into the output measure. However, this feedback is highly frequent, disaggregated, and thus difficult to analyze (Casas-Arce et al., 2017). Even though individual-level sales data has been available for sales employees at the fresh food counter, as sales employees record their sales on their individual number, the company did not analyze or communicate it in the past.

## 4. Experimental Design

### *4.1 The Performance Variable*

The company's IT system captures how much revenue sales employees generate and how many transactions they perform. Based on these key figures, the average sales per transaction per sales employee can be calculated.

$$Average\ Sales\ per\ Transaction = \frac{\sum Sales}{\#Transactions}$$

The performance indicator average sales per transaction is a common metric in retail (Bullard, 2016) and used at our company to analyze sales performance at the store and department level. However, since it reflects how much employees sell per transaction, the average sales per transaction is also an indicator of the performance of sales employees. The average sales per transaction is preferred over alternative measures such as total sales because it is more comparable across employees. Using absolute measures such as total sales would not be appropriate for social comparison as the number of transactions is highly dependent on the number of hours worked and employees differ significantly on the hours worked. The average sales per transaction, however, is comparable across employees independent of the number of customers served.

*4.2 The Treatments*

We introduce a digital app that enables sales employees in the butchery, cheese, and fish departments to analyze their average sales per transaction. Our study's experimental design consists of a no RPI group (*Control*) and two treatment groups (*Median RPI*, *Distribution RPI*). The control and treatment groups each receive different reports on the average sales per transaction in the app.[11] The information is displayed in two-week periods and updated on a weekly basis. A time interval of two weeks instead of, for example, a daily interval avoids random variance, thereby increasing the informativeness of RPI (Hannan et al., 2019).

Although no performance based incentives exist, employees can gain utility from outperforming their peers in the performance reports provided during the experiment. First, they may feel good when receiving positive feedback on their performance report enhancing their self-image (Suls & Wheeler, 2000). Second, they can disclose their performance reports to supervisors and peers voluntarily to achieve social distinction when they perform well (Frey, 2007; Hannan et al., 2013). On the other hand, when they perform poorly, there is no pressure to disclose. Supervisors are not allowed to ask employees for their performance due to the works councils' requirement for anonymity. Moreover, not all employees are expected to use the app. Thus, it is not unusual for an employee not to disclose her report. Since the institutional environment in our setting creates the opportunity to gain utility for employees who outperform their peers but no disutility for employees who underperform their peers, it can be considered generally ahead-seeking.

The different treatments are displayed in Figure 2 and are as follows:

**Control**

Employees in the *Control* Group (the no RPI condition) receive a report about their average sales per transaction. They are informed about their average sales per transaction over the past two weeks as well as the development of their average sales per transaction over the last eight weeks.[12] The main reason for this design choice is to avoid possible contamination effects due to the implementation of the app and effects that are driven simply by increased attention on the new performance metric (Manthei et al., 2023). In order to be able to distinguish the effect

[11] The company prefers such kind of private RPI over public RPI, because it considers it less offensive and thus easier to implement with regard to its works council and data privacy laws.

[12] In our discussions with the company's practitioners, it emerged that providing information about the trend for the last weeks is important to make the app attractive and interesting to employees as it allows them to see whether they are on a positive or negative trend.

of the treatments from the effect of the app's introduction, employees of all groups receive the design of the *Control* group at first.

**Median RPI**

Employees in the *Median RPI* treatment group receive a report similar to the control group. In addition, they receive information about the median average sales per transaction of employees from their department from comparable stores over the last two weeks.[13] Below the chart on the median of the comparison group, employees are informed in a sentence whether their average sales per transaction exceeds the median average sales per transaction of the comparison group. This is framed in such a way that only the potential positive distance to the reference point is highlighted to foster an ahead-seeking environment, i.e. *"You are among the Top 50%. This means that on average you do generate more revenue per sale than 50% of the employees in the comparison group."*[14]

**Distribution RPI**

Employees in the *Distribution RPI* treatment group receive a report similar to the control group. In addition, they receive the average sales per transaction of the performance deciles of employees from their department from comparable stores over the last two weeks. Below the graph on the deciles of the comparison group, employees are informed in a sentence to which decile they belong. Like in the *Median RPI* treatment, this is framed in a way that only the potential positive distance to the other deciles is highlighted (ahead-seeking), i.e. *"You are among the Top 10%. This means that on average you generate more revenue per transaction than 90% of the employees in the comparison group."*

*Insert Figure 2 about here.*

[13] To ensure that RPI is fair for employees, homogeneity among peers is ensured by forming three peer groups based on store size. In other words, each employee receives RPI that is calculated based on similar stores. Peer groups are formed based on store size because it is easy to understand for employees and reasonable as store size determines the assortment employees have available as well as local demand differences of customers. Within-store peer groups would have also been possible and might have increased employees' interest in RPI (Mahlendorf et al., 2014). However, in small stores with few employees, this conflicts with keeping RPI private. Furthermore, forming ten deciles (*Distribution RPI*) requires at least ten employees per department in a store.

[14] For employees who are below the median the text is "You are not among the Top 50%. This means that on average you do not generate more revenue per sale than 50% of the employees in the comparison group."

*4.3 The Implementation*

The experimental intervention lasted for a period of three months, from the beginning of November 2021 to the end of January 2022. The stores were assigned to one of the three groups described above using clustered stratified randomization based on the peer group (store size) and stores' prior average sales per transaction. We clustered our randomization on the store level to avoid contamination effects.[15] The intervention took place in two phases and the timeline of the project is displayed in Figure 3. One month prior the experiment (phase one, October) all sales employees working in the butchery, cheese, and fish departments received information about the newly developed app. We provided this information via a personal letter which also contained employees' login data (see Appendix B 1). The letter was handed over to employees by their store manager.[16] To increase awareness of the app, the information articulated in the letter was also presented to the employees by their store manager. We provided the store manager with a script on how to react to specific questions in advance to ensure that employees receive the same answers. In addition, a DINA3 poster displaying the app was placed at the employees' workplace (see Appendix B 1). This procedure gave us high control over the information processed while raising awareness and participation for the app.

We initiated the experiment (phase two, November to January) by replacing the DINA3 posters in the stores of the treatment groups *Median RPI* and *Distribution RPI*. The new posters informed employees that they would receive information on the relative performance of employees in their department working in comparable stores from now on (see Appendix B 1). To increase awareness of the new poster, store managers informed employees that the poster had changed.[17]

At the end of the experiment, i.e., the beginning of February 2022, we invited employees of all treatment groups to participate in a survey regarding the app via a personal letter handed over by their store manager (see Appendix B 3).[18] To ensure the anonymity of the employees, the survey was implemented by one of the authors' home institution. The survey comprised questions related to employee satisfaction, stress, and social comparison, as well as questions

---

[15] Employees usually do not change stores. Since we randomize at the store level it is thus excluded that employees from one treatment group receive information on their reference group from another treatment group.

[16] Delivering a letter through the store manager is in line with general practice in the organization of interest.

[17] To further improve the usage rate of the app, we sent out a personal reminder letter to employees via their store managers at the start of the first week of January (see Appendix B 2). This letter was the same for all employees regardless of their treatment group with the sole purpose of reminding employees about the app and their login data.

[18] Due to requirements from the work counsel we had to frame the survey and the survey questions so that employees could clearly relate them to the average sales report.

on the usage of the information provided. To increase participation, completion of the survey was incentivized with the possibility of winning one of three shopping vouchers, each worth €100 (approximately 110 USD). Furthermore, to improve participation in the survey, employees from 10 stores were supported with a tablet on which they could conduct the survey and personnel advice by an assistant external to their store after two weeks of running the survey.

To ensure the integrity of the experiment, we implemented in a rigorous procedure that prevented participants from inferring their participation in a university-led study. Store managers, department managers, and employees were not informed that they were part of an experiment. The company used the word "pilot" for internal communication, which is a common practice for firm-level field experiments (Friebel et al., 2017). This enabled us to maintain a natural environment in combination with randomization thereby combining the advantages of control and realism (Floyd & List, 2016).

*Insert Figure 3 about here.*

### *4.4 Data Collection*

Our analysis draws on four distinct sources of data. First, we use data from the company's IT system, which provides data on generated sales and the number of customers served per employee. Second, we utilize data from the company's personnel records containing details on employees' age, gender, length of service, and position. Third, we collect information on the usage rate from the app's reporting system. And fourth, we incorporate information from the questionnaire, which was administered at the end of the intervention.

Our study involves 42 stores situated in southwestern Germany, each featuring a butchery, cheese, and fish department. Employees are included in the analysis when they worked at least four weeks during the experiment. The total number of employees included in the study is 596. Table 1 presents an overview of descriptive statistics on the participating stores, departments and employees.

*Insert Table 1 about here.*

# 5. Results

*5.1 Regression Specification*

To investigate our hypotheses, we run a fixed effects difference-in-difference regression with employee fixed effects as well as time fixed effects. The respective regression equation is:

$$(1)\ Y_{i,t} = \beta_0 + \beta_1 * Median_{i,t} + \beta_2 * Distribution_{i,t} + X_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t}$$

$Y_{i,t}$ denotes the individual average sales per transaction for employees *i* in week *t*. $X_{i,t}$ captures time-variant controls (employees weekly hours worked). $\delta_t$ are weekly time fixed effects, $\alpha_i$ are individual fixed effects of the employee.[19] The error term $\varepsilon_{i,t}$ is clustered on the store level. $Median_{i,t}$ is a dummy variable equal to 1 for employees in the *Median RPI* treatment during the experimental period and 0 otherwise. $Distribution_{i,t}$ is a dummy variable equal to 1 for employees in the *Distribution RPI* treatment during the experimental period and 0 otherwise. Our regression's reference group is the group *Control*, i.e., employees who only received their own performance information. We use data from August 2021 until the end of the experimental period in January 2022. The estimated treatment effects from the above regression give us intention-to-treat effects (ITT), which we compare to the control group and between each other using a Wald test.

We additionally provide estimation results using an ANCOVA regression (McKennzie, 2012). This simple OLS approach has been shown to be a valid robustness check (e.g., Burlig et al., 2020; Friebel et al., 2017). The respective regression equation is:

$$(2)\ Y_{i,t} = \beta_0 + \beta_1 * Median_{i,t} + \beta_2 * Distribution_{i,t} + \overline{Y_{i,t}} + X_{i,t} + \delta_t + \varepsilon_{i,t}$$

The notation is similar to the fixed effects regression. Importantly, we only use data from the experimental period for this estimation approach. Yet, $\overline{Y_{i,t}}$ represents the pre-experimental mean of the dependent variable (the individual average sales per transaction). A list of variables can be found in the appendix (Appendix A, Table A 15).

[19] Note that department and store fixed effects are similar to employee fixed effects, as employees do not switch to other stores or departments during the experimental period.

*5.2 Average Treatment Effect*

To test the average performance effect of our treatments, we use the above regression equation (1) with the individual average sales per transaction in the respective department for employee *i* in week *t* as the dependent variable ($Average\ Sale_{i,t}$). The results are displayed in Table 2, column 1. Furthermore, to check the robustness of our findings, we perform three additional analyses. First, in columns 2 and 3, we perform two OLS (ANVOCA) regressions which control for employees' prior average sales per transaction as well as further employee and store-related control variables (regression equation (2)).[20] Second, in column 4, we use the above regression equation (1) with the average sale per transaction on the store level as the dependent variable. Third, we estimate the local average treatment effect (LATE) instead of only the intention-to-treat (ITT) effect (Appendix A, Table A 13). This yields the performance effect on those employees who complied with the treatment and used the app.

*Insert Table 2 about here.*

With values of -€0.015 and -€0.002, the point estimates for the treatments *Median RPI* and *Distribution RPI* are not statistically significant (Table 2, column 1).[21] Compared to prior performance, this would equal a performance effect of -0.163%, i.e., -0.022%. Comparing the effects of *Median RPI* and *Distribution RPI* using a Wald test does not yield a statistically significant difference between the effects of the treatments (Table 2).

Our robustness checks on the ITT yield similar results (Table 2, columns 2-4). Furthermore, considering the take-up of the treatment, we find that 33.7% (N = 201) of the employees opened their performance report at least once during the experiment (Appendix A, Table A 12).[22] Estimating the LATE among compliers provides effects similar to the ITT (Appendix A, Table A 13).

[20] Control variables used are counter length, tenure, gender, weekly hours worked and dummy variables for the respective weeks as well as dummy variables for the reference group employees were (see footnote 11).

[21] Please note, however, that the respective confidence bands are rather wide. The 95% confidence interval for the point estimate of *Median RPI* is [-0.223; 0.193] (Table 2, column 1). The 95% confidence interval for the point estimate of *Distribution RPI* is [-0.222; 0.219] (Table 2, column 1). Moreover, please note that using wild-cluster bootstrapping does not alter the *p*-values much (Cameron et al., 2008).

[22]This is similar to Eyring and Narayanan (2018) and Kohler et al. (2023), who find that 26.08%, i.e. 30.50% of the employees access their performance report.

Thus, we do not find support for H1 that private RPI has a significant positive average effect on performance.

*5.3 Heterogeneous Effect Depending on Prior Performance*

The above indicates that our treatments do not significantly affect employees' performance on average. However, the absence of an average effect does not imply that RPI does not affect employees at all. Contrarily, theory suggests that RPI has different effects on high performers and low performers (Roels & Su, 2014). In consequence, positive and negative effects may cancel out. Following Roels and Su (2014), our second hypothesis focuses on the effect of median and distribution RPI depending on employees' prior performance. Prior performance is significantly related to employees' performance during the experiment and, thus, regarding our hypotheses, a viable indicator of their expectations to outperform their peers.

To estimate the polarizing effect of RPI with our performance data, we interact the treatment indicator variables with employees' average performance during the 12 weeks prior to the experiment ($Prior\ Performance_i$) as well as employees' estimated fixed effect ($FE\ Absolute_i$) in regression equation (1) (Table 3, column 1 and 4). The fixed effect, $FE\ Absolute_i$, is captured by performing regression equation (3) with weekly data from before the experiment, i.e., August 2021 – October 2021.[23] Precisely, it captures the time-invariant proportion of individual performance that cannot be explained by other variables in the regression, thereby describing differences in employees' prior absolute performance.[24] Furthermore, we use employees' ranked prior performance percentile ($Prior\ Performance\ \ Percentile_i$) and their ranked fixed effect percentile ($FE\ Percentile_i$), both indicating employees' prior relative position in the performance distribution, i.e., to outperform others (Table 3, columns 2 and 5).[25] Finally, we create two dummy variables

[23] (3) $Y_{i,t} = \beta_0 + X_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t}$ , were $Y_{i,t}$ denotes the individual average sales per transaction for employees i in week t. $X_{i,t}$ captures time-variant controls (employees weekly hours worked). $\delta_t$ are weekly time fixed effects, $\alpha_i$ are individual fixed effects of the employee. The error term $\varepsilon_{i,t}$ is clustered on the store level.

[24] As employees usually do not switch between stores, we cannot separate individual performance from store performance. Compared to Bertrand and Schoar (2003), this means we cannot separate 'the manager' fixed effect from 'the firm' fixed effect. However, with regard to our hypotheses, it does not matter why an employee achieves a high performance. For our study, what matters is the performance distance employees see in the performance report.

[25] To be precise, employees' percentile is calculated by dividing an employee's individual rank by the number of all employees such that the variable takes a value of 1 for the employee with the highest performance and takes a value close to 0 for the employees with the lowest performance. It is an interesting robustness check, as it reduces the influence of the performance differences between two employees. See, for instance, Aggrawal and Samwick (1999) and Manthei et al. (2021) for a similar approach.

$Prior\ Performance\ Top\ Quartile_i$ and $FE\ Top\ Quartile_i$, which indicate whether employees' prior performance is in the top quartile (Table 3, columns 3 and 6).

*Insert Table 3 about here.*

The results suggest that there is a polarizing effect, i.e., the performance effect of RPI is more positive for high performers than for low performers (Table 3). Point estimates for the interaction terms of *Median RPI* and *Distribution RPI* and the different indicators of prior performance are positive and economically relevant (Table 3). However, the interaction terms are only statistically significant for the *Distribution RPI* treatment. For instance, the point estimate of the interactions between $Distribution_{i,t}$ and $Prior\ Performance\ Percentile_i$ and $FE\ Percentile_i$, i.e., employees' relative performance percentile, are €0.57 and €0.64 (Table 3, column 2 and 5).[26] Given that $Prior\ Performance\ Percentile_i$ and $FE\ Percentile_i$ range between 0 and 1, this suggests that the performance effect of *Distribution RPI* is around €0.57 to €0.64 more positive for an employee who is at the top of the performance distribution than for an employee at the bottom. Compared to the performance ex-ante the experiment, this corresponds to a performance difference between high and low performers of 6.2%, i.e., 6.9% (Table 1). To estimate the effect of *Distribution PRI* on high performers, we combine the interaction effect with the baseline effect. This shows an estimate of 0.274 (*p*-value 0.144) for $Prior\ Performance\ Percentile_i$ and 0.313 (*p*-value 0.101) for $FE\ Percentile_i$ (Table 3, columns 2 and 5). Although not significant at conventional levels, this suggests that the polarization effect might be driven by a positive effect on employees at the top of the performance distribution.

To analyze the effect on different parts of the performance distribution, we split our sample into quartiles based on $Prior\ Performance_i$ (Table 4, columns 1-4) and $FE\ Absolute_i$ (Table 4, columns 5-8) in regression equation (1). For employees from the top quartile, RPI seems to have a positive effect on performance if RPI is presented in the form of the distribution.[27] Point estimates are economically relevant and statistically significant for *Distribution RPI* with values of €0.38 for the top quartile in prior performance and for the top quartile in the fixed effect

[26] Interacting *Distribution RPI* with dummy variables which indicate whether employees prior performance or fixed effect is in the top tercile instead of the top quartile yields a point estimate of 0.376 (*p*-value = 0.042) for prior performance dummy and 0.380 (*p*-value = 0.049) for the fixed effect dummy (Table A 1).

[27] We also provide this analysis on tertials in the appendix (Appendix, Table A 2).

(Table 4, columns 4 and 8). For *Median RPI*, the point estimates for the top quartile are smaller with €0.16 and €0.20 and statistically insignificant (Table 4, columns 4 and 8). Point estimates for low and medium performers are mostly negative and statistically insignificant for both treatments.

Thus, we do find support for H2 that private RPI in an ahead-seeking environment causes polarization by motivating high performers while demotivating low performers. Yet, contrary to H3, we do not find the polarization effect to be more severe for *Median RPI* than for *Distribution RPI*.

*Insert Table 4 about here.*

*5.4 Robustness Checks*

We perform four robustness checks. First, we perform the interaction analysis and the sample split analysis using the ANCOVA regression (regression equation (2), Appendix A, Table A 3 & Table A 4). This provides support for the above results. The interaction terms between $Distribution_{i,t}$ and $Prior\ Performance\ Percentile_i$, $Prior\ Performance\ Top\ Quartile_i$, $FE\ Absolute_i$, $FE\ Percentile_i$, and $FE\ Top\ Quartile_i$ are statistically and economically significant in both specifications (Appendix A, Table A 3). Similarly, the sample split shows a significant positive point estimate for *Distribution RPI* for employees in the top quartile while point estimates for low and medium performers are insignificant and negative (Table A 4).

Second, while the above discussed heterogeneous effects with respect to the prior performance are robust, we have to caution that we did not randomize the employees' prior performance. Hence, it might be that the displayed analyses are confounded by other variables that are correlated with prior performance and performance during the experiment.[28] In other words, it is possible that while we think we pick up the effect of prior performance, we might pick up a different, yet correlated, effect. To analyze and control for potential cofounds of employees' prior performance, Table 5 includes further interaction terms with other employee and store characteristics. First, column 1 replicates the prior regression (Table 3, column 5).

[28] Importantly, usage rate of the app is not driven by prior performance or other available control variables (Appendix, Table A 11).

Columns 2-7 add the interaction terms age, tenure, store size, counter length, location, and gender separately.[29] Finally, column 8 adds all mentioned interaction terms simultaneously. As visible, our heterogeneous treatment effects with respect to employees' prior performance remain stable and even gain in size and precision when controlling for these variables, increasing confidence in the above results. The same analysis with $FE\ Absolute_i$, $Prior\ Performance_i$, and $Prior\ Performance\ Percentile_i$ as well as with using regression equation (2), is provided in the Appendix (Appendix A, Table A 5, Table A 6, Table A 7, and Table A 8).

*Insert Table 5 about here.*

Third, to address concerns that employees' prior performance may not equal what employees saw in the app, we implement the interaction analysis and the quartile split analysis with prior performance defined as employees' performance during the two weeks prior to the experiment, i.e., the performance employees saw when the treatments were introduced (Appendix A, Table A 9 & Table A 10). Again, all interaction terms between $Distribution_{i,t}$ and $Prior\ Performance_i$, $Prior\ Performance\ Percentile_i$, and $Prior\ Performance\ Top\ Quartile_i$ are positive and statistically significant (Appendix A, Table A 9). Furthermore, similarly to the above analysis, the point estimate of *Distribution RPI* is positive and statistically significant for top performers in the top quartile but statistically insignificant for low and medium performers as well as for *Median RPI* (Appendix A, Table A 10). This again provides support for the above findings.

Fourth, we provide the quartile split estimating the LATE instead of the ITT, i.e. the treatment effect on employees who complied with the treatment and used the app (Appendix A, Table A 14). Similarly, to the above analysis this shows that *Distribution RPI* positively affects employees at the top of the performance distribution while having negative but statistically insignificant point estimates for employees at the bottom or in the middle of the performance distribution. Again, there is no statistically significant effect for *Median RPI*.

Finally, we make use of the described employee survey (N = 120 | 20.1 %) to gain additional evidence on employees` motivation depending on their performance. Specifically, we asked

[29] Location is a dummy variable equaling 1 when the store is located in a major regional center.

employees how much they agree (on a scale from 1 to 7) with the statements: “I feel a great sense of personal satisfaction when I perform well in the average sales report.” and “I feel bad and unhappy when I discover that I have performed poorly on the average sales report.” at the end of the experiment. In line with the above results, employees from all groups state that they indeed feel rather satisfied when performing well than that they feel bad when performing poorly (Figure 4). This difference is statistically significant with a *p*-value of <0.001 using a two-sided Wilcoxon Signed Rank test.[30] This is consistent with the argumentation that our environment can be characterized as an ahead-seeking environment in which employees gain utility from outperforming peers but no disutility from underperforming their peers.

*Insert Figure 4 about here.*

*5.5 Discussion*

As described above, our results suggest that RPI does not significantly affect performance, on average. The absence of an average treatment effect for RPI is in line with Roels and Su (2014) but contrasts H1, i.e., that RPI motivates employees to increase performance, a hypothesis which management accounting studies regularly derive from social comparison theory (e.g., Azmat & Iriberri, 2010; Hannan et al., 2013; Tafkov, 2013). Considering the conflicting results of prior studies testing private RPI in the field (e.g., Ashraf et al., 2014; Blanes i Vidal & Nossol, 2011), our study highlights that private RPI cannot be assumed to have a positive effect on performance on average. Instead, in line with Eyring and Narayanan (2018), our study suggests that one has to consider potential heterogeneous effects of RPI depending on employees’ position in the performance distribution.

Contrary to our second hypothesis, we do not find a statistically significant demotivating effect for *Median RPI* and *Distribution RPI* among low performers. One potential explanation might be that, in our setting, employees are subject to informal controls which hinder low performers from decreasing their performance further. Working side by side with their colleagues and their department manager, they have to serve customers’ requests at the

[30] Note that we pooled all groups together because the number of responses does not allow conclusions about statistical differences between the treatment groups but rather about the general perception of the setting. Figure A 1 in the Appendix displays the mean outcomes for the full survey with confidence bars on a 10% significance level.

minimum. Thus, although private RPI may demotivate low performers, this effect might not materialize in the average sales per transaction. On a more general level, another potential explanation could be that low performers expected to be able to beat their peers. As long as the reference point seems reachable, RPI might not demotivate low performers to give up (Eyring & Narayanan, 2018; Roels & Su, 2014).

Regarding the effect on high performers, our results oppose Roels and Su's (2014) proposition that *Distribution RPI* is more motivating than *Median RPI*. Building on Eyring and Narayanan (2018), who find that the performance effect of RPI decreases the further one outperforms the reference point, we suggest this might be because, in contrast to Roels and Su (2014), the utility gained from outperforming peers is not linear. Furthermore, employees may not assign equal value to each reference point if presented with multiple reference points. For example, Gill et al. (2019) find that being at the top is especially motivating. Thus, we suggest that *Distribution RPI* might have a more positive effect on the performance of high performers than *Median RPI* because it shifts their attention from "beating the mean" to "being at the top".

Considering prior studies testing RPI in a company setting, our results align with Barankay (2012) and Ashraf (2022) regarding the absence of an average performance effect for private RPI.[31] However, in contrast to both studies, we find that private RPI can positively affect the performance of employees at the top of the performance distribution. We suggest that differences in the settings might provide an explanation. In both studies, employees receive 100% of their compensation based on piece rates, while in our study, employees do not receive performance-based pay. Thus, in Barankay (2012) and Ashraf (2022), RPI might be unable to further increase employees' motivation. Considering Blanes i Vidal and Nossol (2011), our study supports the finding that employees at the top of the performance distribution increase their performance as a result of RPI. However, unlike Blanes i Vidal and Nossol (2011), we do not find this effect for employees in the middle and at the bottom of the performance distribution. Again, the differences might be explained by piece rate incentives, which composite 25% of workers' pay in Blanes i Vidal and Nossol (2011). Receiving piece rate incentives, low performers might act on RPI because it helps them improve their performance-based compensation (Tafkov, 2013), not out of social comparison concerns. Considering the

[31] Barankay (2012) finds a negative performance effect for rank RPI. However, this negative effect disappears if ranks are combined with information on absolute performance, i.e. the performance of the top 10%, top 25%, top 50%, which is similar to our RPI design.

absence of explicit incentives and the private digital communication channel, our results might be closer to the isolated effect of social comparison.

## 6. Conclusion

To the best of our knowledge, our study is the first field experiment testing the effect of private RPI on performance against a no RPI condition in a company setting without potentially confounding incentives such as performance-based pay. Specifically, we examine how providing RPI as a median (*Median RPI*) or providing RPI as deciles (*Distribution RPI*) affects performance compared to a no RPI condition (*Control*) depending on employees' position in the performance distribution.

Our study has three main findings. First, we do not find that RPI, on average, affects performance. Second, in line with our hypotheses, we find that this might be explained by heterogeneous effects that cancel each other out. In this regard, our results suggest that *Distribution RPI* risks polarization by motivating employees at the top of the performance distribution but not low and medium performers. Third, in contrast to our hypothesis, the polarization effect does not seem more severe for *Median RPI* than for *Distribution RPI*. If anything, the opposite is the case, as the polarization effect is only significant for *Distribution RPI*.

Considering evidence on the effects of different RPI designs on low and high performers (e.g., Kohler et al., 2023; Rilke et al., 2021), future research may find it valuable to utilize the opportunities of digitalization and test the targeted assignment (e.g., Opitz et al., 2023) of specific RPI designs. For instance, using digital dashboards, one might direct low performers' attention to RPI based on effort and high performers attention to RPI based on performance (Rilke et al., 2021). In this context, potential researchers might evaluate employees' acceptance of such targeted RPI designs. Finally, similar to Kohler et al. (2023), our study highlights that future research should consider subjects' motivation to access their treatments. In this regard, researchers may find it valuable to study how employees can be motivated to look at their performance reports and how firms can foster habits of working with performance reports.

Importantly, on a more general level, our study highlights that RPI can improve performance but that organizations have to be careful about how it affects employees across the performance distribution. High performers seem to benefit from RPI. However, companies should carefully monitor how it affects their low performers.

## References


Aggrawal, R. K., & Samwick, A. A. (1999). The Other Side of the Trade-Off: The Impact of Risk on Executive Compensation. *Journal of Political Economy*, *107*(1), 65-105. https://doi.org/10.1086/250051

Ashraf, A. (2022). Performance Ranks, Conformity, and Cooperation: Evidence from a Sweater Factory. *CESifo Working Paper*, *No. 9591*. https://doi.org/10.2139/ssrn.4042435

Ashraf, N., Bandiera, O., & Lee, S. S. (2014). Awards Unbundled: Evidence from a Natural Field Experiment. *Journal of Economic Behavior and Organization*, *100*, 44-63. https://doi.org/10.1016/j.jebo.2014.01.001

Azmat, G., & Iriberri, N. (2010). The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students. *Journal of Public Economics*, *94*(7-8), 435-452. https://doi.org/10.1016/j.jpubeco.2010.04.001

Banker, R. D., Lee, S.-Y., & Potter, G. (1996). A Field Study of the Impact of a Performance-Based Incentive Plan. *Journal of Accounting and Economics*, *21*(2), 195-226. https://doi.org/doi.org/10.1016/0165-4101(95)00418-1

Barankay, I. (2012). *Rank Incentives: Evidence from a Randomized Workplace Experiment*. Wharton School, University of Pennsylvania. Retrieved 10.02.2023 from http://repository.upenn.edu/bepp_papers/75

Belnap, A. (2023). The Effect of Intermediary Coverage on Disclosure: Evidence From a Randomized Field Experiment. *Journal of Accounting and Economics*, *75*(1).

Bertrand, M., & Schoar, A. (2003). Managing with Style: The Effect of Managers on Firm Policies. *Quarterly Journal of Economics*, *118*(4), 1169-1208. https://doi.org/10.1162/003355303322552775

Blader, S., Gartenberg, C., & Prat, A. (2020). The Contingent Effect of Management Practices. *Review of Economic Studies*, *87*(2), 721-749. https://doi.org/10.1093/restud/rdz034

Blanes i Vidal, J., & Nossol, M. (2011). Tournaments Without Prizes: Evidence from Personnel Records. *Management Science*, *57*(10), 1721-1736. http://www.jstor.org/stable/41261969

Brade, R., Himmler, O., & Jäckle, R. (2023). Relative Performance Feedback and Long-Term Tasks - Experimental Evidence From Higher Education. *CESifo Working Paper*, *No. 10346*. https://doi.org/10.2139/ssrn.4410248

Bullard, B. (2016). *Style and Statistics: The Art of Retail Analytics*. The Wiley and SAS Business Series.

Burlig, F., Louis, P., & Woerman, M. (2020). Panel Data and Experimental Design. *Journal of Development Economics*, *144*(C), 1-15. https://doi.org/10.3386/w26250

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference With Clustered Errors. *The Review of Economics and Statistics*, *90*(3), 414-427. https://doi.org/10.1162/rest.90.3.414

Casas-Arce, P., Lourenço, S. M., & Martínez-Jerez, F. A. (2017). The Performance Effect of Feedback Frequency and Detail: Evidence from a Field Experiment in Customer Satisfaction. *Journal of Accounting Research*, *55*(5), 1051-1088. https://doi.org/10.1111/1475-679X.12184

Costello, A. M., Down, A. K., & Mehta, M. N. (2020). Machine + Man: A field Experiment on the Role of Discretion in Augmenting AI-Based Lending Models. *Journal of Accounting and Economics*, *70*.

Delfgaauw, J., Dur, R., Sol, J., & Verbeke, W. (2013). Tournament Incentives in the Field: Gender Differences in the Workplace. *Journal of Labor Economics*, *31*(2), 305-326. https://doi.org/10.1086/667996

Eyring, H. (2020). Disclosing Physician Ratings: Performance Effects and the Difficulty of Altering Ratings Consensus. *Journal of Accounting Research*, *58*(4), 1023-1067. https://doi.org/10.1111/1475-679X.12330

Eyring, H., & Narayanan, V. G. (2018). Performance Effects of Setting a High Reference Point for Peer-Performance Comparison. *Journal of Accounting Research*, *56*(2), 581-615. https://doi.org/10.1111/1475-679X.12199

Festinger, L. (1954). A Theory of Social Comparison. *Human Relations*, *7*, 117-140.

Fletcher, T. D., & Nusbaum, D. N. (2010). Development of the Competitive Work Environment Scale: A Multidimensional Climate Construct. *Educational and Psychological Measurement*, *70*(1), 105-124. https://doi.org/10.1177/0013164409344492

Floyd, E., & List, J. A. (2016). Using Field Experiments in Accounting and Finance. *Journal of Accounting Research*, *54*(2), 437-475. https://doi.org/10.1111/1475-679X.12113

Frey, B. S. (2007). Awards as Compensation. *European Accounting Review*, *4*(1), 6-14. https://doi.org/10.1057/palgrave.emr.1500068

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. https://doi.org/10.1257/aer.20160788

Gilbert, P., Broomhead, C., Irons, C., McEwan, K., Bellew, R., Mills, A., Gale, C., & Knibb, R. (2007). Development of a striving to avoid inferiority scale. *British Journal of Social Psychology*, *46*(3), 633-648. https://doi.org/10.1348/014466606x157789

Gill, D., Kissová, Z., Lee, J., & Prowse, V. (2019). First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision. *Management Science*, *65*(2), 494-507. https://doi.org/10.1287/mnsc.2017.2907

Hannan, R. L., Krishnan, R., & Newman, A. H. (2008). The Effects of Disseminating Relative Performance Feedback in Tournament and Individual Performance Compensation Plans. *The Accounting Review*, *83*(4), 893-913. https://doi.org/10.2308/accr.2008.83.4.893

Hannan, R. L., McPhee, G. P., Newman, A. H., & Tafkov, I. D. (2013). The Effect of Relative Performance Information on Performance and Effort Allocation in a Multi-Task

Environment. *The Accounting Review*, *88*(2), 553-575. https://doi.org/10.2308/accr-50312

Hannan, R. L., McPhee, G. P., Newman, A. H., & Tafkov, I. D. (2019). The Informativeness of Relative Performance Information and its Effect on Fffort Allocation in a Multitask Environment. *Contemporary Accounting Research*, *36*(3), 1607-1633. https://doi.org/10.1111/1911-3846.12482

Hermes, H., Huschens, M., Rothlauf, F., & Schunk, D. (2021). Motivating Low-Achievers - Relative Performance Feedback in Primary Schools. *Journal of Economic Behavior and Organization*, *187*, 45-59. https://doi.org/10.1016/j.jebo.2021.04.004

Kelly, K., Presslee, A., & Webb, R. A. (2017). The Effects of Tangible Rewards Versus Cash Rewards in Consecutive Sales Tournaments: A Field Experiment. *The Accounting Review*, *92*(6), 165-185. https://doi.org/10.2308/accr-51709

Kohler, M., Mahlendorf, M. D., Seiter, M., & Vogelsang, T. (2023). *Social Comparison on Multiple Tasks - Sacrificing Overall Performance for Local Excellence.* [Unpublished Registered Report].

Kramer, S., Maas, V. S., & Rinsum, M. v. (2016). Relative Performance Information, Rank Ordering, and Employee Performance: A Research Note. *Management Accounting Research*, *33*, 16-24. https://doi.org/10.1016/j.mar.2016.03.004

Kuhnen, C. M., & Tymula, A. (2012). Feedback, Self-Esteem, and Performance in Organizations. *Management Science*, *58*(1), 94-113. https://doi.org/10.1287/mnsc.1110.1379

Lazear, E. P. (2000). Performance Pay and Productivity. *American Economic Review*, *90*(5), 1346-1361. https://doi.org/10.1257/aer.90.5.1346

Lourenço, S. M. (2016). Monetary Incentives, Feedback, and Recognition - Complements or Substitutes? Evidence from a Field Experiment in a Retail Services Company. *The Accounting Review*, *91*(1), 279-297. https://doi.org/10.2308/accr-51148

MacDonald, S., & MacIntyre, P. (1997). The Generic Job Satisfaction Scale: Scale Development and its Correlates. *Employee Assistance Quarterly*, *13*(2), 1-16. https://doi.org/10.1300/J022v13n02_01

Mahlendorf, M. D., Kleinschmit, F., & Perego, P. (2014). Relational Effects of Relative Performance Information: The Role of Professional Identity. *Accounting, Organizations and Society*, *39*(5), 331-347. https://doi.org/10.1016/j.aos.2014.05.001

Manthei, K., Sliwka, D., & Vogelsang, T. (2021). Performance Pay and Prior Learning - Evidence from a Retail Chain. *Management Science*, *67*(11), 6998-7022. https://doi.org/10.1287/mnsc.2020.3825

Manthei, K., Sliwka, D., & Vogelsang, T. (2023). Information, Incentives, and Attention: A Field Experiment on the Interaction of Management Controls. *The Accounting Review*, *98*(5), 455-479. https://doi.org/10.2308/TAR-2019-1065

McKennzie, D. (2012). Beyond Baseline and Follow-Up: The Case for More T in Experiments. *Journal of Development Economics*, *99*(2), 210-221. https://doi.org/10.1016/j.jdeveco.2012.01.002

Newby, J. L., & Klein, R. G. (2014). Competitiveness Reconceptualized: Psychometric Development of the Competitiveness Orientation Measure as a Unified Measure of Trait Competitiveness. *Psychological Record*, *64*(4), 879-895. https://doi.org/10.1007/s40732-014-0083-2

Opitz, S., Sliwka, D., Vogelsang, T., & Zimmermann, T. (2023). The Algorithmic Assignment of Incentive Schemes. *Management Science*(forthcoming).

Rilke, R. M., van Pelt, V., Lehnen, S., & Guenther, C. (2021). Motivating Low Performers with Input-based Relative Performance Feedback. *SSRN*, *December*(16). https://doi.org/10.2139/ssrn.3978948

Roels, G., & Su, X. (2014). Optimal Design of Social Comparison Effects: Setting Reference Groups and Reference Points. *Management Science*, *60*(3), 606-627. https://doi.org/10.1287/mnsc.2013.1760

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychol Sci*, *18*(5), 429-434. https://doi.org/10.1111/j.1467-9280.2007.01917.x

Smith, R. H. (2000). Assimilative and Contrastive Emotional Reactions to Upward and Downward Social Comparisons. In J. Suls & L. W. Wang (Eds.), *Handbook of Social Comparison: Theory and Research* (pp. 173-200). Kluwer Academic Publishers. https://doi.org/10.1007/978-1-4615-4237-7_10

Song, H., Tucker, A. L., Murrell, K. L., & Vinsonc, D. R. (2018). Closing the Productivity Gap: Improving Worker Productivity Through Public Relative Performance Feedback and Validation of Best Practices. *Management Science*, *64*(6), 2628-2649. https://doi.org/10.1287/mnsc.2017.2745

Spreitzer, G. M. (1995). Psychological Emporwerment in the Workplace: Dimensions, Measurement, and Validation. *The Academy of Management Journal*, *38*(5), 1442-1465. https://doi.org/10.2307/256865

Suls, J., & Wheeler, L. (Eds.). (2000). *Handbook of Social Comparison - Theory and Research*. Kluwer Academic Publishers. https://doi.org/10.1007/978-1-4615-4237-7.

Tafkov, I. D. (2013). Private and Public Relative Performance Information under Different Compensation Contracts. *The Accounting Review*, *88*(1), 327-350. https://doi.org/10.2308/accr-50292

Tran, A., & Zeckhauser, R. (2012). Rank as an Inherent Incentive: Evidence From a Field Experiment. *Journal of Public Economics*, *96*(9-10), 645-650. https://doi.org/10.1016/j.jpubeco.2012.05.004

# Tables and Figures

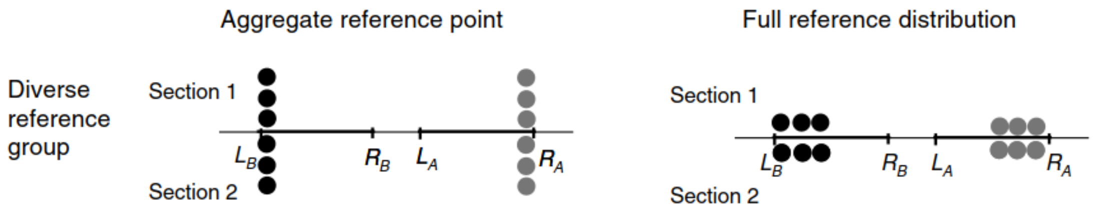



Figure 1: Effect of RPI in an Ahead-seeking Environment (Roels and Su 2014, p. 614).


Figure 2: The Treatments

*Note:* The figure shows the different treatment. The text in the treatment Median RPI reads “You are among the Top 50%. This means that on average you do generate more revenue per sale than 50% of the employees in the comparison group.”. The text in treatment Distribution RPI reads “You are among the Top 10%. This means that on average you generate more revenue per transaction than 90% of the employees in the comparison group.”.

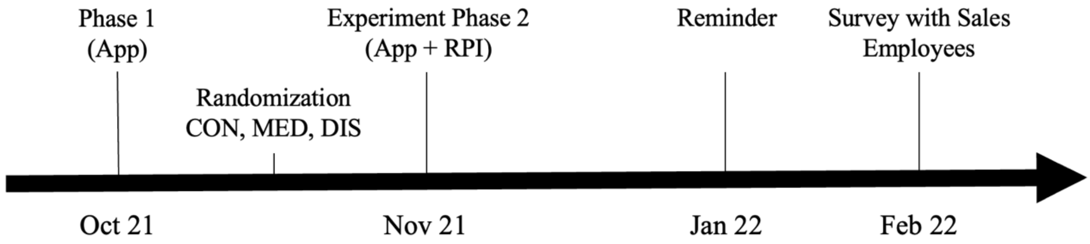



Figure 3: Timeline

Table 1: Descriptive Statistics

| | *N* | Mean | Std. Dev. | 25% | 75% |
|---|---|---|---|---|---|
| Female Employee (1/0) | 596 | 0.81 | 0.39 | 1 | 1 |
| Age | 596 | 44.05 | 13.30 | 33 | 47 |
| Tenure | 596 | 10.31 | 9.66 | 3 | 16 |
| Store Size ($m^2$) | 42 | 3,086 | 1,265 | 2,275 | 3,801 |
| Length Fresh Food Counters (m) | 42 | 25.55 | 6.58 | 20 | 26 |
| Weekly Sales per Employee Own Department in € | 566 | 2,210.45 | 1,257.52 | 1,185.00 | 3,079.51 |
| Average Sale per Transaction Own Department in € | 566 | 9.22 | 1.81 | 8.21 | 9.93 |

*Note:* The table reports descriptive statistics for the overall sample. The sales data is winsorized at the 1% and 99% percentile. The sales data covers a period of 12 weeks prior to October 25, 2021. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in the department in which he worked. This results in 596 employees being included in the experiment. Of these 596 employees, 566 have been active during the 12 weeks prior to the experiment.

Table 2: Main Treatment Effects

| | (1) Average Sale | (2) Average Sale | (3) Average Sale | (4) Average Sale |
|---|---|---|---|---|
| *Median RPI* | -0.015 | -0.030 | -0.031 | 0.053 |
| | (0.103) | (0.122) | (0.090) | (0.072) |
| *Distribution RPI* | -0.002 | -0.005 | -0.023 | 0.050 |
| | (0.109) | (0.136) | (0.115) | (0.070) |
| Wald test ($p$ – value) | 0.877 | 0.796 | 0.485 | 0.970 |
| Time FE | Yes | No | No | Yes |
| Individual FE | Yes | No | No | No |
| Store FE | No | No | No | Yes |
| Average Sale Prior | No | Yes | Yes | No |
| Controls | Yes | No | Yes | No |
| Unit of Observation | Employee | Employee | Employee | Store |
| N of Observations | 12,251 | 6,271 | 6,271 | 1,078 |
| N of Employees | 596 | 566 | 566 | 596 |
| N of Stores | 42 | 42 | 42 | 42 |
| $R^2$ | 0.592 | 0.485 | 0.497 | 0.926 |

*Note:* The table reports results from a fixed effects regression with the average sales per transaction on the (1) employee and (4) store level as the dependent variable. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Furthermore, the table reports results from ANCOVA regressions with the average sales per transaction on the employee level as the dependent variable. The regressions account for employees' performance during the 12 weeks prior the experiment (2) as well as calendar week, tenure, age, store size, region, department, weekly hours worked, gender, counter length, and reference group (3). The regressions are built on observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to a linear regression estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. The Wald test tests for equality among the point estimates of *Median RPI* and *Distribution RPI*. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Impact of Prior Performance on Treatment Effect

| | (1) Average Sale | (2) Average Sale | (3) Average Sale | (4) Average Sale | (5) Average Sale | (6) Average Sale |
|---|---|---|---|---|---|---|
| *Median RPI* | 0.009<br>(0.092) | -0.189<br>(0.186) | 0.065<br>(0.120) | 0.010<br>(0.094) | -0.219<br>(0.204) | -0.085<br>(0.124) |
| *Distribution RPI* | 0.011<br>(0.103) | -0.295<br>(0.193) | -0.145<br>(0.125) | 0.010<br>(0.103) | -0.324<br>(0.204) | -0.149<br>(0.125) |
| *Median RPI* ×<br>Prior Performance | 0.131<br>(0.105) | | | | | |
| *Distribution RPI* ×<br>Prior Performance | 0.186<br>(0.112) | | | | | |
| *Median RPI* ×<br>Prior Performance Percentile | | 0.369<br>(0.276) | | | | |
| *Distribution RPI* ×<br>Prior Performance Percentile | | 0.569*<br>(0.305) | | | | |
| *Median RPI* ×<br>Prior Performance Top Quartile | | | 0.225<br>(0.234) | | | |
| *Distribution RPI* ×<br>Prior Performance Top Quartile | | | 0.526**<br>(0.247) | | | |
| *Median RPI* ×<br>FE Absolute | | | | 0.121<br>(0.092) | | |
| *Distribution RPI* ×<br>FE Absolute | | | | 0.175*<br>(0.098) | | |
| *Median RPI* ×<br>FE Percentile | | | | | 0.439<br>(0.305) | |
| *Distribution RPI* ×<br>FE Percentile | | | | | 0.638*<br>(0.324) | |
| *Median RPI* ×<br>FE Top Quartile | | | | | | 0.287<br>(0.217) |
| *Distribution RPI* ×<br>FE Top Quartile | | | | | | 0.530**<br>(0.243) |
| Wald test<br>(*p* – value) | 0.307 | 0.424 | 0.058 | 0.337 | 0.433 | 0.157 |
| Time FE ×<br>Interaction Variable | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 11,971 | 11,971 | 11,971 | 11,964 | 11,964 | 11,964 |
| N of Employees | 566 | 566 | 566 | 565 | 565 | 565 |
| $R^2$ | 0.613 | 0.611 | 0.609 | 0.612 | 0.613 | 0.610 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. $Prior\ Performance$ equals employees' performance during the 12 weeks prior to the experiment and is mean centered. $Prior\ Performance\ Percentile$ equals the percentile of employee's prior performance. $Prior\ Performance\ Top\ Quartile$ is a dummy variable that takes a value of 1 if employees' prior performance is in the top quartile and 0 if it is not. $FE\ Absolute$ equals employees' absolute fixed effect based on data prior to the experiment. The variable is mean centered and indicates whether they are able to achieve high average sales per transaction. $FE\ Percentile$ equals the percentile of employees' fixed effect. $FE\ Top\ Quartile$ is a dummy variable that takes a value of 1 if employees' fixed effect is in the top quartile and 0 if it is not. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. The Wald test tests for equality among the point estimates of *Median RPI* and Interaction Variable and *Distribution RPI* and Interaction Variable. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Impact of Prior Performance on Treatment Effect, Quartile Split

| | Prior Performance | | | | FE Absolute | | | |
|---|---|---|---|---|---|---|---|---|
| | 1st Quartile | 2nd Quartile | 3rd Quartile | 4th Quartile | 1st Quartile | 2nd Quartile | 3rd Quartile | 4th Quartile |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale |
| *Distribution RPI* | -0.059<br>(0.162) | -0.166<br>(0.144) | -0.146<br>(0.179) | 0.382*<br>(0.213) | -0.082<br>(0.159) | -0.215<br>(0.157) | -0.082<br>(0.147) | 0.381*<br>(0.212) |
| Wald test<br>(*p* – value) | 0.334 | 0.021 | 0.708 | 0.213 | 0.521 | 0.130 | 0.455 | 0.338 |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 2,921 | 3,051 | 3,006 | 2,993 | 2,926 | 3,024 | 3,038 | 2,976 |
| N of Employees | 141 | 142 | 141 | 142 | 141 | 141 | 141 | 142 |
| $R^2$ | 0.395 | 0.301 | 0.338 | 0.590 | 0.395 | 0.283 | 0.380 | 0.558 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. The sample is split in quartiles based on employees' performance during the 12 weeks prior to the experiment (1-4) and employees' fixed effect $FE\ Absolute$ (5-8). $FE\ Absolute$ equals employees' absolute fixed effect and indicates whether they are able to achieve high average sales per transaction. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Interaction Effect with FE Percentile and Different Control Variables

| | (1) Average Sale | (2) Average Sale | (3) Average Sale | (4) Average Sale | (5) Average Sale | (6) Average Sale | (7) Average Sale | (8) Average Sale |
|---|---|---|---|---|---|---|---|---|
| *Median RPI* | -0.219<br>(0.204) | -0.493<br>(0.503) | -0.293<br>(0.208) | -0.184<br>(0.222) | -0.812**<br>(0.385) | -0.247<br>(0.184) | -0.197<br>(0.194) | -0.198<br>(0.378) |
| *Distribution RPI* | -0.324<br>(0.204) | -0.284<br>(0.518) | -0.294<br>(0.230) | -0.135<br>(0.253) | -0.190<br>(0.338) | -0.285<br>(0.183) | -0.310<br>(0.194) | 0.342<br>(0.383) |
| *Median RPI* × FE Percentile | 0.439<br>(0.305) | 0.441<br>(0.312) | 0.419<br>(0.306) | 0.600*<br>(0.308) | 0.549*<br>(0.306) | 0.523<br>(0.364) | 0.465<br>(0.349) | 0.770*<br>(0.456) |
| *Distribution RPI* × FE Percentile | 0.638*<br>(0.324) | 0.650*<br>(0.335) | 0.631*<br>(0.323) | 0.771**<br>(0.321) | 0.639*<br>(0.321) | 0.720*<br>(0.378) | 0.722*<br>(0.358) | 1.042**<br>(0.443) |
| *Median RPI* × Age | | 0.006<br>(0.008) | | | | | | 0.004<br>(0.008) |
| *Distribution RPI* × Age | | -0.001<br>(0.009) | | | | | | -0.002<br>(0.008) |
| *Median RPI* × Tenure | | | 0.008<br>(0.006) | | | | | -0.001<br>(0.009) |
| *Distribution RPI* × Tenure | | | -0.003<br>(0.007) | | | | | -0.007<br>(0.010) |
| *Median RPI* × Size | | | | -0.000<br>(0.000) | | | | -0.000*<br>(0.000) |
| *Distribution RPI* × Size | | | | -0.000<br>(0.000) | | | | -0.000<br>(0.000) |
| *Median RPI* × Length | | | | | 0.020*<br>(0.011) | | | 0.001<br>(0.010) |
| *Distribution RPI* × Length | | | | | -0.005<br>(0.009) | | | -0.013*<br>(0.008) |
| *Median RPI* × Location | | | | | | -0.008<br>(0.227) | | -0.112<br>(0.261) |
| *Distribution RPI* × Location | | | | | | -0.108<br>(0.245) | | -0.346<br>(0.270) |
| *Median RPI* × Gender (1 = Male) | | | | | | | -0.223<br>(0.380) | -0.367<br>(0.393) |
| *Distribution RPI* × Gender (1 = Male) | | | | | | | -0.336<br>(0.406) | -0.508<br>(0.422) |
| Time FE × Interaction Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 11,964 | 11,964 | 11,964 | 11,964 | 11,964 | 11,964 | 11,964 | 11,964 |
| N of Employees | 565 | 565 | 565 | 565 | 565 | 565 | 565 | 565 |
| $R^2$ | 0.613 | 0.614 | 0.615 | 0.614 | 0.614 | 0.614 | 0.615 | 0.622 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. The treatment indicator variables are interacted with the variables $FE\ Percentile$, $Age$, $Tenure$, $Size$, $Length$, $Location$, and $Gender$. $FE\ Percentile$ indicates the rank of employees' fixed effect. $Size$ describes store size in square meter. $Length$ describes length of the fresh food counter in meter. $Location$ is a dummy variable and equals 1 when the store is operating in a major regional center. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

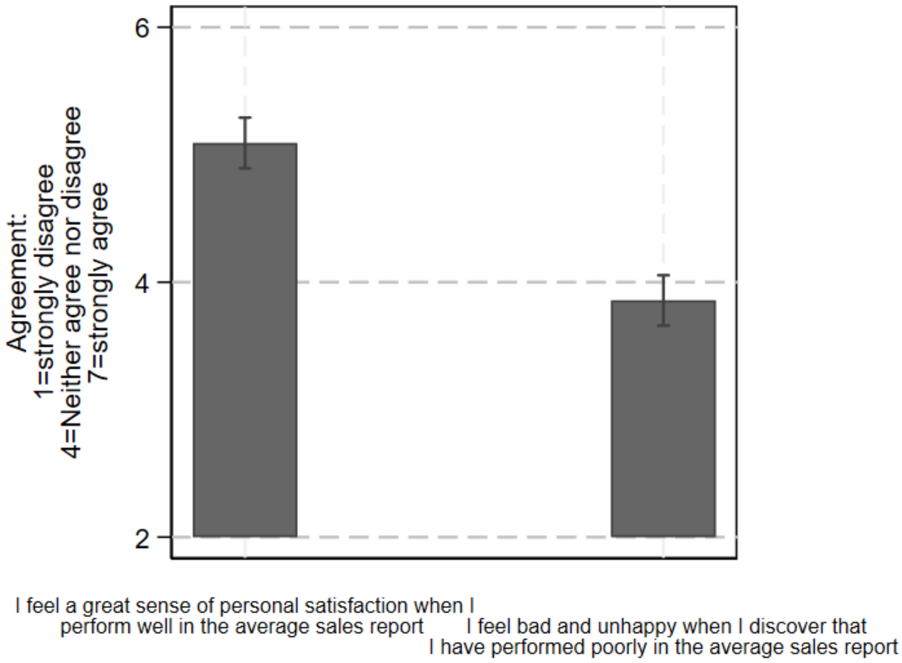



Figure 4: Effect of Feedback on Satisfaction

*Note*: This figure shows the average answer of employees to Q6-Q7 for employees of different treatment groups (1=strongly disagree, 7=strongly agree). Q6 "I feel a great sense of personal satisfaction when I perform well in the average sales report". Q7 "I feel bad and unhappy when I discover that I have performed poorly in the average sales report". 90% confidence bars are displayed. N = 120.

## Appendix A (Tables and Figures)

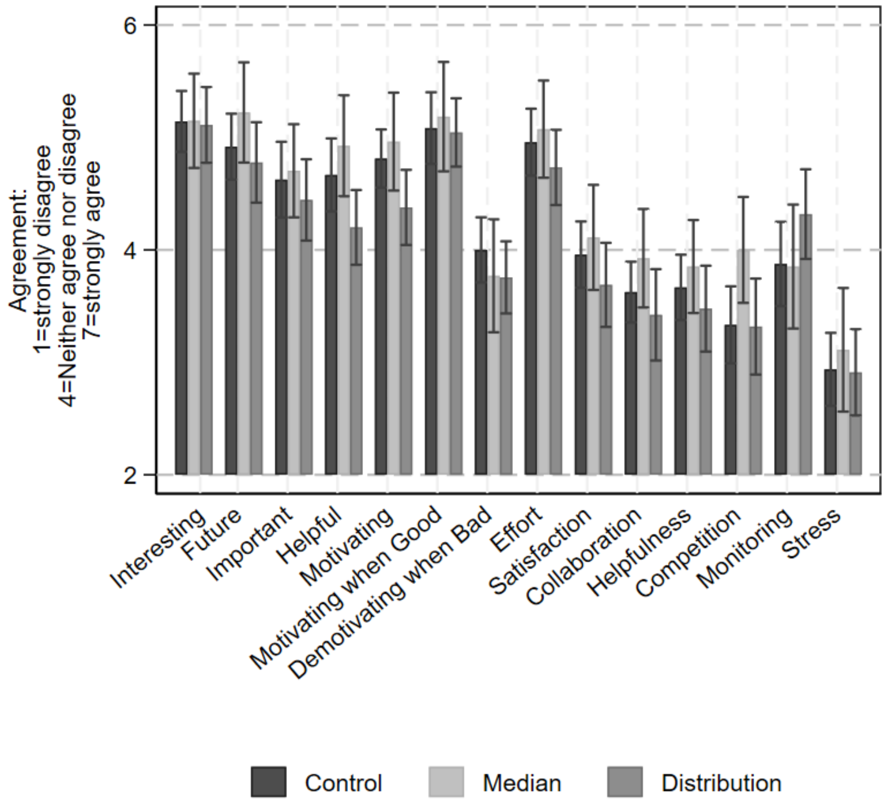



Figure A 1: The Survey

*Note*: This figure shows the average answer of employees to Q1-Q14 (left to right) for different treatment groups (1=strongly disagree, 7=strongly agree). 90% confidence bars are displayed. N = 122.

Table A 1: Impact of Prior Performance on Treatment Effect Tertial

| | (1)<br>Average Sale | (2)<br>Average Sale |
|---|---|---|
| *Median RPI* | -0.046<br>(0.108) | -0.041<br>(0.107) |
| *Distribution RPI* | -0.144<br>(0.115) | -0.144<br>(0.115) |
| *Median RPI* ×<br>Prior Performance Top Tertial | 0.082<br>(0.154) | |
| *Distribution RPI* ×<br>Prior Performance Top Tertial | 0.376**<br>(0.179) | |
| *Median RPI* ×<br>FE Top Tertial | | 0.078<br>(0.157) |
| *Distribution RPI* ×<br>FE Top Tertial | | 0.380**<br>(0.187) |
| Wald test<br>($p$ – value) | 0.384 | 0.048 |
| Time FE ×<br>Interaction Variable | Yes | Yes |
| Time FE | Yes | Yes |
| Individual FE | Yes | Yes |
| Controls | Yes | Yes |
| Department | All | All |
| Unit of Observation | Employee | Employee |
| N of Observations | 11,971 | 11,964 |
| N of Employees | 566 | 565 |
| $R^2$ | 0.611 | 0.611 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. $Prior\ Performance$ equals employees' performance during the 12 weeks prior to the experiment and is mean centered. $Prior\ Performance\ Top\ Tertial$ is a dummy variable that takes a value of 1 if employees' prior performance is in the top tertial and 0 if it is not. $FE\ Top\ Tertial$ is a dummy variable that takes a value of 1 if employees' fixed effect is in the top tertial and 0 if it is not. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. The Wald test tests for equality among the point estimates of *Median RPI* and Interaction Variable and *Distribution RPI* and Interaction Variable. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 2: Impact of Prior Performance on Treatment Effect, Tertial Split

| | Prior Performance | | | FE Absolute | | |
|---|---|---|---|---|---|---|
| | 1st Tercile | 2nd Tercile | 3rd Tercile | 1st Tercile | 2nd Tercile | 3rd Tercile |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale |
| *Median RPI* | -0.078 | 0.009 | 0.038 | -0.112 | 0.048 | 0.040 |
| | (0.158) | (0.127) | (0.156) | (0.157) | (0.124) | (0.158) |
| *Distribution RPI* | -0.083 | -0.188 | 0.234 | -0.084 | -0.177 | 0.238 |
| | (0.141) | (0.133) | (0.176) | (0.140) | (0.120) | (0.182) |
| Wald test ($p$ – value) | 0.963 | 0.098 | 0.178 | 0.794 | 0.059 | 0.200 |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 3,917 | 4,020 | 4,034 | 3,911 | 4,055 | 3,998 |
| N of Employees | 188 | 188 | 190 | 188 | 188 | 189 |
| $R^2$ | 0.401 | 0.340 | 0.559 | 0.408 | 0.344 | 0.549 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. The sample is split in tertials based on employees' performance during the 12 weeks prior to the experiment (1-3) and employees fixed effect $FE\ Absolute$ (4-6). $FE\ Absolute$ equals employees' absolute fixed effect and indicates whether they are able to achieve high average sales per transaction and is mean centered. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 3: Impact of Prior Performance on Treatment Effect (ANCOVA)

| | (1) Average Sale | (2) Average Sale | (3) Average Sale | (4) Average Sale | (5) Average Sale | (6) Average Sale |
|---|---|---|---|---|---|---|
| *Median RPI* | 0.029 | -0.332 | -0.050 | -0.003 | -0.379 | -0.092 |
| | (0.086) | (0.278) | (0.105) | (0.082) | (0.278) | (0.104) |
| *Distribution RPI* | 0.049 | -0.434 | -0.147 | -0.009 | -0.515 | -0.165 |
| | (0.092) | (0.324) | (0.137) | (0.095) | (0.313) | (0.135) |
| *Median RPI* × Prior Performance | 0.245 | | | | | |
| | (0.195) | | | | | |
| *Distribution RPI* × Prior Performance | 0.304 | | | | | |
| | (0.194) | | | | | |
| *Median RPI* × Prior Performance Percentile | | 0.650 | | | | |
| | | (0.499) | | | | |
| *Distribution RPI* × Prior Performance Percentile | | 0.873* | | | | |
| | | (0.511) | | | | |
| *Median RPI* × Prior Performance Top Quartile | | | 0.164 | | | |
| | | | (0.240) | | | |
| *Distribution RPI* × Prior Performance Top Quartile | | | 0.562** | | | |
| | | | (0.240) | | | |
| *Median RPI* × FE Absolute | | | | 0.188 | | |
| | | | | (0.174) | | |
| *Distribution RPI* × FE Absolute | | | | 0.259* | | |
| | | | | (0.146) | | |
| *Median RPI* × FE Percentile | | | | | 0.755 | |
| | | | | | (0.500) | |
| *Distribution RPI* × FE Percentile | | | | | 1.047** | |
| | | | | | (0.500) | |
| *Median RPI* × FE Top Quartile | | | | | | 0.309 |
| | | | | | | (0.228) |
| *Distribution RPI* × FE Top Quartile | | | | | | 0.669** |
| | | | | | | (0.248) |
| Wald test ($p$ – value) | 0.287 | 0.374 | 0.020 | 0.472 | 0.194 | 0.041 |
| Time FE × Interaction Variable | No | No | No | No | No | No |
| Time FE | No | No | No | No | No | No |
| Individual FE | No | No | No | No | No | No |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 6,271 | 6,271 | 6,271 | 6,265 | 6,265 | 6,265 |
| N of Employees | 566 | 566 | 566 | 565 | 565 | 565 |
| $R^2$ | 0.501 | 0.498 | 0.498 | 0.507 | 0.500 | 0.499 |

*Note:* The table reports results from an ANCOVA regression the average sales per transaction on the employee level as the dependent variable. The regressions account for employees' performance during the 12 weeks prior the experiment, calendar week, tenure, age, store size, region, department, weekly hours worked, gender, counter length, and reference group. $Prior\ Performance$ equals employees' performance during the 12 weeks prior to the experiment and is mean centered. $Prior\ Performance\ Percentile$ equals the percentile of employees' prior performance. $Prior\ Performance\ Top\ Quartile$ is a dummy variable that takes a value of 1 if employees' prior performance is in the top quartile and 0 if it is not. $FE\ Absolute$ equals employees' absolute fixed effect based on data prior to the experiment. The variable is mean centered and indicates whether they are able to achieve high average sales per transaction. $FE\ Percentile$ indicates the percentile of employees' fixed effect. $FE\ Top\ Quartile$is a dummy variable that takes a value of 1 if employees' fixed effect is in the top quartile and 0 if it is not. The regressions are built on observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to a linear regression estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. The Wald test tests for equality among the point estimates of *Median RPI and* Interaction Variable as well as *Distribution RPI* and Interaction Variable. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 4: Impact of Prior Performance on Treatment Effect, Quartile Split (ANCOVA)

| | Prior Performance | | | | FE Absolute | | | |
|---|---|---|---|---|---|---|---|---|
| | 1st Quartile | 2nd Quartile | 3rd Quartile | 4th Quartile | 1st Quartile | 2nd Quartile | 3rd Quartile | 4th Quartile |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale |
| *Median RPI* | -0.228<br>(0.240) | 0.020<br>(0.117) | -0.015<br>(0.152) | 0.120<br>(0.238) | -0.247<br>(0.256) | -0.098<br>(0.138) | -0.083<br>(0.128) | 0.220<br>(0.220) |
| *Distribution RPI* | -0.336<br>(0.273) | -0.053<br>(0.128) | -0.066<br>(0.132) | 0.394**<br>(0.162) | -0.437<br>(0.303) | -0.175<br>(0.151) | -0.084<br>(0.129) | 0.453***<br>(0.159) |
| Wald test<br>($p$ – value) | 0.668 | 0.562 | 0.630 | 0.155 | 0.500 | 0.515 | 0.998 | 0.238 |
| Time FE | No | No | No | No | No | No | No | No |
| Individual FE | No | No | No | No | No | No | No | No |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 1,555 | 1,591 | 1,572 | 1,553 | 1,558 | 1,579 | 1,581 | 1,547 |
| N of Employees | 141 | 142 | 142 | 142 | 141 | 141 | 141 | 142 |
| $R^2$ | 0.183 | 0.280 | 0.358 | 0.538 | 0.192 | 0.261 | 0.392 | 0.526 |

*Note:* The table reports results from ANCOVA regressions with the employees' average sales per transaction as the dependent variable. The sample is split in quartiles based on employees' performance during the 12 weeks prior to the experiment (1-4) and employees fixed effect $FE\ Absolute$ (5-8). $FE\ Absolute$ equals employees' absolute fixed effect and indicates whether they are able to achieve high average sales per transaction. The regressions are built on observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to a linear regression estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. The Wald test tests for equality among the point estimates of *Median RPI and Distribution RPI*. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 5 : Interaction Effect with FE Percentile and Different Control Variables (ANCOVA)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale | Average Sale |
| *Median RPI* | -0.379<br>(0.278) | -1.035<br>(0.749) | -0.505<br>(0.302) | -0.133<br>(0.244) | -0.693*<br>(0.409) | -0.365<br>(0.242) | -0.377<br>(0.272) | -1.119*<br>(0.598) |
| *Distribution RPI* | -0.515<br>(0.313) | -0.584<br>(0.744) | -0.536<br>(0.364) | -0.122<br>(0.277) | -0.277<br>(0.346) | -0.399*<br>(0.230) | -0.518*<br>(0.299) | 0.237<br>(0.456) |
| *Median RPI* × FE Percentile | 0.755<br>(0.500) | 0.774<br>(0.515) | 0.747<br>(0.496) | 0.862<br>(0.547) | 0.805<br>(0.493) | 0.860<br>(0.583) | 0.883<br>(0.654) | 1.209<br>(0.793) |
| *Distribution RPI* × FE Percentile | 1.047**<br>(0.500) | 1.096**<br>(0.506) | 1.036**<br>(0.501) | 1.182**<br>(0.556) | 1.054**<br>(0.505) | 1.175*<br>(0.592) | 1.269*<br>(0.646) | 1.594*<br>(0.798) |
| *Median RPI* × Age | | 0.015<br>(0.012) | | | | | | 0.012<br>(0.010) |
| *Distribution RPI* × Age | | 0.001<br>(0.011) | | | | | | -0.004<br>(0.008) |
| *Median RPI* × Tenure | | | 0.012<br>(0.008) | | | | | 0.002<br>(0.009) |
| *Distribution RPI* × Tenure | | | 0.002<br>(0.008) | | | | | -0.003<br>(0.010) |
| *Median RPI* × Size | | | | -0.000<br>(0.000) | | | | -0.000**<br>(0.000) |
| *Distribution RPI* × Size | | | | -0.000<br>(0.000) | | | | -0.000<br>(0.000) |
| *Median RPI* × Length | | | | | 0.012<br>(0.014) | | | 0.026<br>(0.017) |
| *Distribution RPI* × Length | | | | | -0.010<br>(0.008) | | | -0.010<br>(0.008) |
| *Median RPI* × Location | | | | | | -0.079<br>(0.288) | | -0.202<br>(0.344) |
| *Distribution RPI* × Location | | | | | | -0.413<br>(0.315) | | -0.472<br>(0.343) |
| *Median RPI* × Gender (1 = Male) | | | | | | | -0.384<br>(0.554) | -0.432<br>(0.551) |
| *Distribution RPI* × Gender (1 = Male) | | | | | | | -0.586<br>(0.587) | -0.717<br>(0.589) |
| Time FE × Interaction Variables | No | No | No | No | No | No | No | No |
| Time FE | No | No | No | No | No | No | No | No |
| Individual FE | No | No | No | No | No | No | No | No |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 6,265 | 6,265 | 6,265 | 6,265 | 6,265 | 6,265 | 6,265 | 6,265 |
| N of Employees | 565 | 565 | 565 | 565 | 565 | 565 | 565 | 565 |
| $R^2$ | 0.500 | 0.501 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.503 |

*Note:* The table reports results from an ANCOVA regression the average sales per transaction on the employee level as the dependent variable. The regressions account for employees' performance during the 12 weeks prior the experiment, calendar week, tenure, age, store size, region, department, weekly hours worked, gender, counter length, and reference group. $FE\ Percentile$ indicates the rank of employees' fixed effect. $Size$ describes store size in square meter. $Length$ describes length of the fresh food counter in meter. $Location$ is a dummy variable and equals 1 when the store is operating in a major regional center. The regressions are built on observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to a linear regression estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 6: Interaction Effect with FE Absolute and Different Control Variables

| | (1)<br>Average Sale | (2)<br>Average Sale | (3)<br>Average Sale | (4)<br>Average Sale | (5)<br>Average Sale | (6)<br>Average Sale | (7)<br>Average Sale | (8)<br>Average Sale |
|---|---|---|---|---|---|---|---|---|
| *Median RPI* | 0.010 | -0.289 | -0.072 | 0.194 | -0.459 | 0.030 | 0.058 | 0.354 |
| | (0.094) | (0.407) | (0.107) | (0.186) | (0.340) | (0.167) | (0.080) | (0.393) |
| *Distribution RPI* | 0.010 | 0.045 | 0.039 | 0.332 | 0.164 | 0.098 | 0.077 | 1.039$^{**}$ |
| | (0.103) | (0.430) | (0.128) | (0.203) | (0.295) | (0.171) | (0.077) | (0.430) |
| *Median RPI* × FE Absolute | 0.121 | 0.120 | 0.119 | 0.165 | 0.136 | 0.137 | 0.138 | 0.220 |
| | (0.092) | (0.093) | (0.092) | (0.100) | (0.094) | (0.106) | (0.105) | (0.134) |
| *Distribution RPI* × FE Absolute | 0.175$^{*}$ | 0.177$^{*}$ | 0.174$^{*}$ | 0.213$^{**}$ | 0.178$^{*}$ | 0.192$^{*}$ | 0.197$^{*}$ | 0.280$^{**}$ |
| | (0.098) | (0.100) | (0.097) | (0.104) | (0.098) | (0.111) | (0.109) | (0.134) |
| *Median RPI* × Age | | 0.007 | | | | | | 0.005 |
| | | (0.008) | | | | | | (0.007) |
| *Distribution RPI* × Age | | -0.001 | | | | | | -0.002 |
| | | (0.008) | | | | | | (0.007) |
| *Median RPI* × Tenure | | | 0.007 | | | | | -0.003 |
| | | | (0.005) | | | | | (0.009) |
| *Distribution RPI* × Tenure | | | -0.003 | | | | | -0.009 |
| | | | (0.007) | | | | | (0.010) |
| *Median RPI* × Size | | | | -0.000 | | | | -0.000$^{*}$ |
| | | | | (0.000) | | | | (0.000) |
| *Distribution RPI* × Size | | | | -0.000 | | | | -0.000 |
| | | | | (0.000) | | | | (0.000) |
| *Median RPI* × Length | | | | | 0.018 | | | -0.001 |
| | | | | | (0.012) | | | (0.010) |
| *Distribution RPI* × Length | | | | | -0.006 | | | -0.014$^{*}$ |
| | | | | | (0.009) | | | (0.008) |
| *Median RPI* × Location | | | | | | -0.016 | | -0.126 |
| | | | | | | (0.237) | | (0.259) |
| *Distribution RPI* × Location | | | | | | -0.118 | | -0.369 |
| | | | | | | (0.254) | | (0.270) |
| *Median RPI* × Gender (1 = Male) | | | | | | | -0.281 | -0.458 |
| | | | | | | | (0.403) | (0.422) |
| *Distribution RPI* × Gender (1 = Male) | | | | | | | -0.388 | -0.598 |
| | | | | | | | (0.429) | (0.449) |
| Time FE × Interaction Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 11,964 | 11,964 | 11,964 | 11,964 | 11,964 | 11,964 | 11,964 | 11,964 |
| N of Employees | 565 | 565 | 565 | 565 | 565 | 565 | 565 | 565 |
| $R^2$ | 0.612 | 0.613 | 0.615 | 0.614 | 0.614 | 0.613 | 0.615 | 0.623 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. $FE\ Absolute$ indicates employees' fixed effect and is mean centered. $Size$ describes store size in square meter. $Length$ describes length of the fresh food counter in meter. $Location$ is a dummy variable and equals 1 when the store is operating in a major regional center. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. Robust standard errors are clustered on the store level and displayed in parentheses. [c] Variables are mean centered. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 7: Interaction Effect with Prior Performance and Different Control Variables

| | (1) Average Sale | (2) Average Sale | (3) Average Sale | (4) Average Sale | (5) Average Sale | (6) Average Sale | (7) Average Sale | (8) Average Sale |
|---|---|---|---|---|---|---|---|---|
| *Median RPI* | 0.010<br>(0.092) | -0.306<br>(0.412) | -0.071<br>(0.105) | 0.199<br>(0.190) | -0.459<br>(0.343) | 0.031<br>(0.168) | 0.062<br>(0.082) | 0.364<br>(0.405) |
| *Distribution RPI* | 0.011<br>(0.103) | 0.041<br>(0.433) | 0.041<br>(0.126) | 0.338<br>(0.210) | 0.169<br>(0.297) | 0.103<br>(0.174) | 0.079<br>(0.080) | 1.056** <br>(0.441) |
| *Median RPI* × Prior Performance | 0.131<br>(0.105) | 0.130<br>(0.107) | 0.130<br>(0.105) | 0.174<br>(0.114) | 0.147<br>(0.108) | 0.147<br>(0.119) | 0.151<br>(0.119) | 0.233<br>(0.147) |
| *Distribution RPI* × Prior Performance | 0.186<br>(0.112) | 0.189<br>(0.115) | 0.185<br>(0.111) | 0.222* <br>(0.119) | 0.189<br>(0.113) | 0.203<br>(0.126) | 0.207<br>(0.125) | 0.289* <br>(0.151) |
| *Median RPI* × Age | | 0.007<br>(0.008) | | | | | | 0.005<br>(0.007) |
| *Distribution RPI* × Age | | -0.001<br>(0.009) | | | | | | -0.002<br>(0.007) |
| *Median RPI* × Tenure | | | 0.007<br>(0.005) | | | | | -0.003<br>(0.010) |
| *Distribution RPI* × Tenure | | | -0.003<br>(0.007) | | | | | -0.009<br>(0.010) |
| *Median RPI* × Size | | | | -0.000<br>(0.000) | | | | -0.000* <br>(0.000) |
| *Distribution RPI* × Size | | | | -0.000<br>(0.000) | | | | -0.000<br>(0.000) |
| *Median RPI* × Length | | | | | 0.018<br>(0.012) | | | -0.001<br>(0.010) |
| *Distribution RPI* × Length | | | | | -0.006<br>(0.010) | | | -0.014* <br>(0.008) |
| *Median RPI* × Location | | | | | | -0.021<br>(0.242) | | -0.131<br>(0.261) |
| *Distribution RPI* × Location | | | | | | -0.125<br>(0.260) | | -0.375<br>(0.276) |
| *Median RPI* × Gender (1 = Male) | | | | | | | -0.308<br>(0.417) | -0.495<br>(0.438) |
| *Distribution RPI* × Gender (1 = Male) | | | | | | | -0.395<br>(0.445) | -0.613<br>(0.470) |
| Time FE × Interaction Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 11,971 | 11,971 | 11,971 | 11,971 | 11,971 | 11,971 | 11,971 | 11,971 |
| N of Employees | 566 | 566 | 566 | 566 | 566 | 566 | 566 | 566 |
| $R^2$ | 0.613 | 0.614 | 0.615 | 0.615 | 0.614 | 0.614 | 0.616 | 0.624 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. $Prior\ Performance$ describes employees' performance during the 12 weeks prior to the experiment and is mean centered. $Size$ describes store size in square meter. $Length$ describes length of the fresh food counter in meter. $Location$ is a dummy variable and equals 1 when the store is operating in a major regional center. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 8: Interaction Effect with Prior Performance Percentile and Different Control Variables

| | (1) Average Sale | (2) Average Sale | (3) Average Sale | (4) Average Sale | (5) Average Sale | (6) Average Sale | (7) Average Sale | (8) Average Sale |
|---|---|---|---|---|---|---|---|---|
| *Median RPI* | -0.189 | -0.468 | -0.262 | -0.173 | -0.796** | -0.223 | -0.172 | -0.227 |
| | (0.186) | (0.496) | (0.194) | (0.218) | (0.379) | (0.175) | (0.180) | (0.385) |
| *Distribution RPI* | -0.295 | -0.255 | -0.262 | -0.123 | -0.167 | -0.260 | -0.284 | 0.338 |
| | (0.193) | (0.512) | (0.219) | (0.252) | (0.324) | (0.179) | (0.187) | (0.382) |
| *Median RPI* × Prior Performance Percentile | 0.369 | 0.369 | 0.353 | 0.528* | 0.489* | 0.446 | 0.398 | 0.699 |
| | (0.276) | (0.286) | (0.275) | (0.279) | (0.279) | (0.327) | (0.319) | (0.427) |
| *Distribution RPI* × Prior Performance Percentile | 0.569* | 0.584* | 0.564* | 0.694** | 0.569* | 0.643* | 0.654* | 0.957** |
| | (0.305) | (0.317) | (0.303) | (0.300) | (0.303) | (0.348) | (0.342) | (0.420) |
| *Median RPI* × Age | | 0.006 | | | | | | 0.005 |
| | | (0.008) | | | | | | (0.008) |
| *Distribution RPI* × Age | | -0.001 | | | | | | -0.002 |
| | | (0.009) | | | | | | (0.008) |
| *Median RPI* × Tenure | | | 0.007 | | | | | -0.002 |
| | | | (0.006) | | | | | (0.009) |
| *Distribution RPI* × Tenure | | | -0.003 | | | | | -0.008 |
| | | | (0.007) | | | | | (0.010) |
| *Median RPI* × Size | | | | -0.000 | | | | -0.000 |
| | | | | (0.000) | | | | (0.000) |
| *Distribution RPI* × Size | | | | -0.000 | | | | -0.000 |
| | | | | (0.000) | | | | (0.000) |
| *Median RPI* × Length | | | | | 0.020* | | | 0.001 |
| | | | | | (0.011) | | | (0.010) |
| *Distribution RPI* × Length | | | | | -0.005 | | | -0.013* |
| | | | | | (0.009) | | | (0.007) |
| *Median RPI* × Location | | | | | | 0.005 | | -0.098 |
| | | | | | | (0.220) | | (0.256) |
| *Distribution RPI* × Location | | | | | | -0.095 | | -0.333 |
| | | | | | | (0.239) | | (0.266) |
| *Median RPI* × Gender (1 = Male) | | | | | | | -0.211 | -0.353 |
| | | | | | | | (0.381) | (0.396) |
| *Distribution RPI* × Gender (1 = Male) | | | | | | | -0.326 | -0.493 |
| | | | | | | | (0.408) | (0.424) |
| Time FE × Interaction Variables | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Department | All | All | All | All | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee | Employee | Employee | Employee | Employee |
| N of Observations | 11,971 | 11,971 | 11,971 | 11,971 | 11,971 | 11,971 | 11,971 | 11,971 |
| N of Employees | 566 | 566 | 566 | 566 | 566 | 566 | 566 | 566 |
| $R^2$ | 0.611 | 0.612 | 0.614 | 0.613 | 0.613 | 0.612 | 0.613 | 0.621 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. $Prior\ Performance\ Rank$ indicates the percentile of employees' prior performance during the 12 weeks prior to the experiment. $Size$ describes store size in square meter. $Length$ describes length of the fresh food counter in meter. $Location$ is a dummy variable and equals 1 when the store is operating in a major regional center. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 9: Impact of Initial Prior Performance on Treatment Effect

| | (1) Average Sale | (2) Average Sale | (3) Average Sale |
|---|---|---|---|
| *Median RPI* | 0.029<br>(0.080) | -0.040<br>(0.140) | 0.010<br>(0.082) |
| *Distribution RPI* | 0.018<br>(0.085) | -0.187<br>(0.145) | -0.076<br>(0.088) |
| *Median RPI* ×<br>Prior Performance | 0.033<br>(0.036) | | |
| *Distribution RPI* ×<br>Prior Performance | 0.075**<br>(0.033) | | |
| *Median RPI* ×<br>Prior Performance Percentile | | 0.159<br>(0.249) | |
| *Distribution RPI* ×<br>Prior Performance Percentile | | 0.450*<br>(0.253) | |
| *Median RPI* ×<br>Prior Performance Top Quartile | | | 0.147<br>(0.216) |
| *Distribution RPI* ×<br>Prior Performance Top Quartile | | | 0.448*<br>(0.227) |
| Wald test<br>($p$ – value) | 0.295 | 0.246 | 0.059 |
| Time FE ×<br>Interaction Variable | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes |
| Department | All | All | All |
| Unit of Observation | Employee | Employee | Employee |
| N of Observations | 11,557 | 11,557 | 11,520 |
| N of Employees | 545 | 545 | 540 |
| $R^2$ | 0.630 | 0.628 | 0.634 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. The treatment indicator variables are interacted with the variables $Prior\ Performance,$ $Prior\ Performance\ Percentile,$ and $Prior\ Performance\ Top\ Quartile.$ $Prior\ Performance$ equals employees' performance during the two weeks prior to the experiment and is mean centered. $Prior\ Performance\ Percentile$ equals the percentile of employees' prior performance. $Prior\ Performance\ Top\ Quartile$ is a dummy variable that takes a value of 1 if employees' prior performance is in the top quartile and 0 if it is not. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. The Wald test tests for equality among the point estimates of *Median RPI and* Interaction Variable and *Distribution RPI* and Interaction Variable. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 10: Impact of Initial Prior Performance on Treatment Effect, Quartile Split Initial

| | Prior Performance | | | |
|---|---|---|---|---|
| | 1st Quartile | 2nd Quartile | 3rd Quartile | 4th Quartile |
| | (1) | (2) | (3) | (4) |
| | Average Sale | Average Sale | Average Sale | Average Sale |
| *Median RPI* | -0.106 | 0.012 | 0.019 | -0.001 |
| | (0.142) | (0.151) | (0.166) | (0.187) |
| *Distribution RPI* | -0.057 | -0.176 | -0.094 | 0.342* |
| | (0.140) | (0.168) | (0.154) | (0.202) |
| Wald test ($p$ – value) | 0.669 | 0.131 | 0.278 | 0.066 |
| Time FE | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| Department | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee |
| N of Observations | 2,938 | 3,286 | 3,382 | 2,570 |
| N of Employees | 141 | 154 | 161 | 122 |
| $R^2$ | 0.589 | 0.520 | 0.491 | 0.560 |

*Note:* The table reports results from a fixed effects regression with the employees' average sales per transaction as the dependent variable. The sample is split in quartiles based on employees' performance during the two weeks prior to the experiment (1-3). The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimators thus refer to the difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 11 – Drivers of Usage Rate OLS

| | (1) Active Once | (2) Active Twice | (3) Active N |
|---|---|---|---|
| Prior Performance | -0.006 | 0.003 | -0.010 |
| | (0.011) | (0.009) | (0.048) |
| Weekly Hours Worked | 0.003 | 0.001 | 0.013 |
| | (0.003) | (0.002) | (0.012) |
| Gender | -0.004 | 0.028 | 0.326 |
| | (0.066) | (0.050) | (0.329) |
| Age | -0.002 | 0.000 | 0.013** |
| | (0.002) | (0.001) | (0.006) |
| Tenure | 0.003 | 0.002 | 0.018 |
| | (0.002) | (0.002) | (0.013) |
| Counter Length in m | -0.000 | -0.001 | 0.002 |
| | (0.004) | (0.003) | (0.011) |
| Store Size in $m^2$ | -0.000 | -0.000 | -0.000 |
| | (0.000) | (0.000) | (0.000) |
| N of Observations | 566 | 566 | 566 |
| $R^2$ | 0.017 | 0.008 | 0.028 |

*Note:* The table reports results from unplanned OLS regressions with the Variables *Active Once*, *Active Twice* and *Active N* as dependent variables. *Active Once* is a dummy variable equaling 1 when an employee opened her report at least once during the experimental period and 0 otherwise. *Active Twice* is a dummy variable equaling 1 when an employee opened her report at least twice during the experimental period and 0 otherwise. *Active N* describes how often an employee accessed his report during the experimental period. The regressions account for prior performance, treatment, weekly hours worked, gender, age, tenure, counter length, and store size. $Prior\ Performance_i$ equals employees' performance during the 12 weeks prior to the experiment. The data is collapsed on the employee level. Observations are excluded when an employee had less than four weeks of sales data in which he worked. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 12: Compliers vs. Non-Compliers

| | (1) ALL | (2) Non-Compliers | (3) Compliers (n>=1) | (4) Compliers (n>=2) |
|---|---|---|---|---|
| Female Employee (1/0) | 0.809 (0.394) | 0.805 (0.397) | 0.816 (0.389) | 0.787 (0.412) |
| Age | 44.054 (13.304) | 44.152 (13.812) | 43.861 (12.274) | 45.382 (11.222) |
| Tenure | 10.312 (9.659) | 10.084 (9.484) | 10.761 (10.004) | 11.989* (10.111) |
| Average Sale | 9.219 (1.810) | 9.242 (1.990) | 9.176 (1.405) | 9.343 (1.530) |
| N of Employees[+] | 596 | 395 | 201 | 89 |
| N of Employees % | 1 | 0.663 | 0.337 | 0.149 |

*Note:* The table reports means of the respective variables for the overall sample with standard deviations in parentheses. The sales data is winsorized at the 1% and 99% percentile. The data covers a period of 12 weeks prior November 2021. T-test Compliers to Non-Compliers. [+]The number of employees is slightly lower for the $Average\ Sale$ as not all employees included in the analysis were active during the 12 weeks prior to the experiment. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 13: Local Average Treatment Effect (LATE)

| | Average Sale<br>Own Department |
|---|---|
| *Median RPI* | -0.056<br>(0.382) |
| *Distribution RPI* | -0.003<br>(0.233) |
| Wald test<br>($p$ – value) | 0.854 |
| Time FE | Yes |
| Individual FE | Yes |
| Controls | Yes |
| Unit of Observation | Employee |
| N of Observations | 12,251 |
| N of Employees | 596 |
| $R^2$ | 0.004 |

*Note:* The table reports results from an instrumental variable regression with fixed effects with the employees' average sales per transaction as the dependent variable. The instrument is $Active\ Once\ During$, a dummy variable indicating whether employees used the app at least once during the experiment. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimator thus refers to the instrumental variable difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. The Wald test tests for equality among the point estimates of *Separate RPI and Separate & Overall RPI*. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 14: Impact of Initial Prior Performance on Treatment Effect (LATE), Quartile Split

| | Prior Performance | | | |
|---|---|---|---|---|
| | 1st Quartile | 2nd Quartile | 3rd Quartile | 4th Quartile |
| | (1) | (2) | (3) | (4) |
| | Average Sale | Average Sale | Average Sale | Average Sale |
| *Median RPI* | -0.655<br>(0.738) | -0.079<br>(0.493) | -0.079<br>(0.493) | 0.922<br>(0.854) |
| *Distribution RPI* | -0.183<br>(0.365) | -0.570<br>(0.448) | -0.570<br>(0.448) | 0.781*<br>(0.449) |
| Wald test<br>($p$ – value) | 0.393 | 0.241 | 0.241 | 0.848 |
| Time FE | Yes | Yes | Yes | Yes |
| Individual FE | Yes | Yes | Yes | Yes |
| Controls | Yes | Yes | Yes | Yes |
| Department | All | All | All | All |
| Unit of Observation | Employee | Employee | Employee | Employee |
| N of Observations | 2,926 | 3,024 | 3,024 | 2,976 |
| N of Employees | 141 | 141 | 141 | 142 |
| $R^2$ | -0.002 | 0.007 | 0.007 | -0.008 |

*Note:* The table reports results from an instrumental variable regression with fixed effects with the employees' average sales per transaction as the dependent variable. The sample is split based on employees' fixed effect $FE\ Absolute$. $FE\ Absolute$ equals employees' absolute fixed effect and indicates whether they are able to achieve high average sales per transaction. The instrument is $Active\ Once\ During$, a dummy variable indicating whether employees used the app at least once during the experiment. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (Aug 2021 – Oct 2021) with the observations during the experiment (Nov 2021 – Jan 2022). The treatment estimator thus refers to the instrumental variable difference-in-difference estimator. Observations are excluded when an employee was absent in the respective week or when an employee had less than four weeks of sales data in which he worked. The Wald test tests for equality among the point estimates of *Separate RPI and Separate & Overall RPI*. Robust standard errors are clustered on the store level and displayed in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A 15: List of Variables

| Variable | Description |
|---|---|
| Active N | Describes how often an employee opened her report during the experimental period. |
| Active Once | A dummy variable equal to 1 when an employee opened the report at least once during the experimental period and 0 otherwise. |
| Active Twice | A dummy variable equal to 1 when an employee opened the report at least twice during the experimental period and 0 otherwise. |
| Age | Describes employees' age in years. |
| Average Sale | The average sales generated by an employee per transaction. |
| Complier (n>=1) | A dummy variable equal to 1 when an employee opened the report at least once during the experimental period and 0 otherwise. |
| Complier (n>=2) | A dummy variable equal to 1 when an employee opened the report at least twice during the experimental period and 0 otherwise. |
| Distribution RPI | A dummy variable equal to 1 for employees in the *Distribution RPI* treatment during the experimental period and 0 otherwise. |
| FE Absolute | Employee fixed effect estimated prior to the experiment using time fixed effects and weekly hours worked as explanatory variables. |
| FE Percentile | The percentile of employees' fixed effect prior to the experiment. |
| FE Top Quartile | A dummy variable equal to 1 for employees whose FE Absolute is in the fourth quartile. |
| FE Top Tercile | A dummy variable equal to 1 for employees whose FE Absolute is in the third tercile. |
| Gender | A dummy variable equal to one if an employee is female. |
| Individual FE | Employees' individual fixed effect based on the data considered in the analysis. |
| Length | The length of the fresh food counters in meter. |
| Location | A dummy variable equal to 1 when the store is located in a major regional center and 0 otherwise. |
| Median RPI | A dummy variable equal to 1 for employees in the *Median RPI* treatment during the experimental period and 0 otherwise. |
| Non-Complier | A dummy variable equal to 1 when an employee did not open the report during the experimental period. |
| Prior Performance | Equals employees' mean centered performance during the 12 weeks prior to the experiment. |
| Prior Performance Percentile | Equals the percentile of employees' performance prior to the experiment. |
| Prior Performance Top Quartile | A dummy variable equal to 1 for employees whose prior performance is in the top quartile. |
| Prior Performance Top Tercile | A dummy variable equal to 1 for employees whose prior performance is in the top tercile. |
| Size | Store size in square meters. |
| Tenure | Employees' tenure with the company in years. |
| Time FE | The fixed effect of calendar weeks. |
| Weekly Hours Worked | Describes the weekly hours worked by employees. |

# Appendix B Online

*Appendix B 1: Kick-Off*

**CONFIDENTIAL**

[redacted]

**Department:** [redacted]

**Phone:** [redacted]

E Ma : [redacted]

Contact: [redacted]

Date: [redacted]

**Pilot Project – My Average Sales**

Dear [redacted]

The average sales per transaction is an important key figure for [redacted]. It shows how much revenue is generated on average per sale. As part of the pilot project „My Average Sales “, you will be able to view **your personal average sales per transaction for the last few weeks** from now on. The aim of the project is to enable you to obtain anonymous feedback on your sales performance via app.

Using the QR code / link, you can access **your personal average sales report**:

**Your Average Sales Report**

Link: [redacted]

QR-Code:

Your report is update once a week, no later than Tuesday morning at 7:00 am. Within the report you have access to your average sales per transaction in one of the following departments: (1) Butchery, (2) Cheese or (3) Fish. Your assignment to a department is based on the sales you generated in recent weeks.

Your personal average sales report is **anonymous** and available to you alone. Your managers will not be informed about your individual average sales during the course of the project. Your individual average sales will not be used anywhere in the company to evaluate your performance. The implementation of the project and the confidential handling of your data has been agreed in advance with your works council.

If you are **assigned to the wrong department** in the report, you can request to be assigned to a different department **till the end of October** at the **following link**. If you have any further questions, you can always contact your store management or [redacted] (see letterhead).

**Request new department assignment**

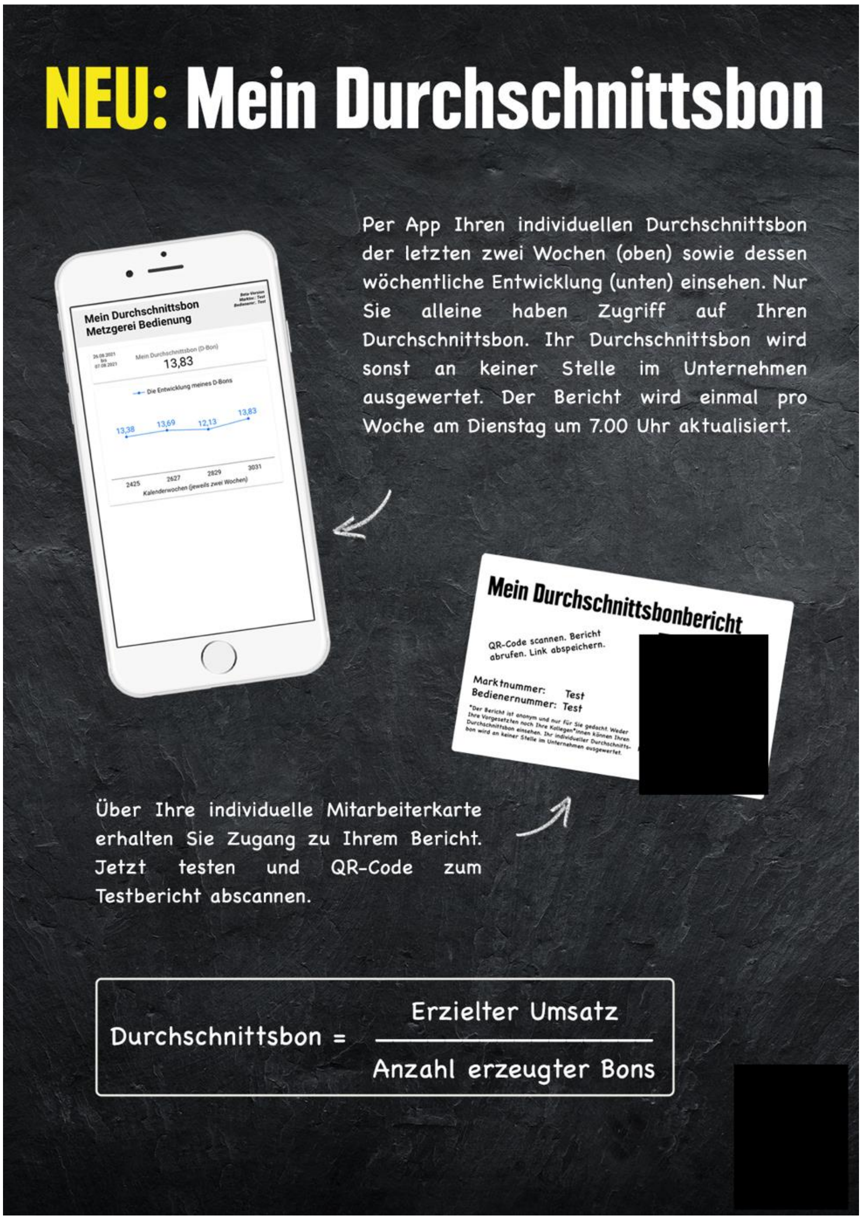
Link: [redacted]

QR-Code:

We are already looking forward to the joint project and would like to thank you in advance for your participation.

Kind regards,

[redacted]




*Note:* Title "New: My Average Sales Report". Text: "Access your individual average sales per transaction during the last two weeks (top) as well as the development over the last weeks (bottom). You alone have access to your average sales report. Your average sales per transaction is not evaluated by anyone in the organization. The report is refreshed once per week on Tuesday 7.00 am. Your individual card provides access data to your report. Test it by scanning the QR code displayed on the card." Formula: "$\frac{Sales}{Number\ of\ Transactions}$"


*Note:* Title "New: Anonymously Benchmark Average Sales". Text: "Using the 'My Average Sales App' you can anonymously compare your average sales with the average sales of comparable peers (Top 50%). There are three peer groups based on store size. Your report displays the average sales of comparable peers in your department. The value Top 50% describes the average sales which at least 50% of comparable employees achieved. Example: The employee 'Me' in the example (left) achieved average sales of €12.13. 50% of employees working in the same department in comparable stores achieved average sales of at least €8.86.


*Note:* Title "New: Anonymously Benchmark Average Sales". Text: "Using the 'My Average Sales App' you can anonymously compare your average sales with the average sales of comparable peers (Top 10%, 20%, …, 100%). There are three peer groups based on store size. Your report displays the average sales of comparable peers in your department. The values Top 10%, 20%, …, 100% describe how many comparable peers achieved average sales of at least €X.XX. Example: The employee 'Me' in the example (left) achieved average sales of €12.13. 10% of employees working in the same department in comparable stores achieved average sales of at least €11.01, 20% average sales of at least €10.43, … .

*Appendix B 2: The Reminder*

**CONFIDENTIAL**
[redacted]

**Department:** [redacted]
**Phone:** [redacted]
E Ma : [redacted]
Contact: [redacted]
Date: [redacted]

Dear [redacted]

a few months have passed since the start of the pilot project „My Average Sales Report". Your feedback as part of the project helps to optimize the provision of information and adapt it to your needs. As we have heard from various sides, the report is already being used intensively. We would like to express our **warmest appreciation for this**.


You are interested in your average sales report over Christmas and New Year's Eve? As usual, you can access your personal average sales report via your individual QR code.

**Your individual access data** (please retain):

[redacted]

Detailed information on the project can be found on the back of the latter. If you have any further questions, please feel free to contact Mr. [redacted] (see letterhead) at any time.

We are looking forward to further developing the report together with you.

Kind regards and a good start into the new year,

[redacted]

**How do I get access to my report?**

**1) Scan QR code**


**2) Open link**

**3) Done**


It is not necessary to download an app. The report opens in your browser (e.g. Safari or Chrome). A login is also not required, as your QR code is unique.

**Background information on the project**

The average sales per transaction is an important key figure for [redacted]. It shows how much revenue is generated on average per sale. As part of the pilot project „My Average Sales ", you will be able to view **your personal average sales per transaction for the last few weeks** from now on. The aim of the project is to enable you to obtain anonymous feedback on your sales performance via app.

Your report is update once a week, no later than Tuesday morning at 7:00 am. Within the report you have access to your average sales per transaction in one of the following departments: (1) Butchery, (2) Cheese or (3) Fish. Your assignment to a department is based on the sales you generated in recent weeks.

Your personal average sales report is **anonymous** and available to you alone. Your managers will not be informed about your individual average sales during the course of the project. Your individual average sales will not be used anywhere in the company to evaluate your performance. The implementation of the project and the confidential handling of your data has been agreed in advance with your works council.

*Appendix B 3: Invitation Survey*

[redacted]

**Survey on the Average Sales Report**

Win one of three €100 shopping vouchers.

[redacted]

[redacted] [redacted]

**Your contacts for the survey**

[redacted]

[redacted]

ppa. [redacted]

[redacted]

[redacted]

**Prof. Dr.** [redacted]

[redacted]

Dear [redacted]

herewith we would like to invite you to participate in a short survey. The survey covers the following topics: The relevance of the average sales report, the usage of the average sales report, and the impact of the average sales report on team spirit and satisfaction.

**Please participate in the survey until 13.02.2022. Your opinion is very important to us! To thank you for your effort, three shopping vouchers worth €100 each will be awarded among survey participants.**

The survey is conducted by the [redacted], which as an independent institution ensures that the questionnaires remain **absolutely anonymous**. Apart from the [redacted], no one will gain insight into the questionnaires. [redacted] **will later only receive average values aggregated over all stores.**

You can participate in the survey via the following QR code / link:

[redacted]

[redacted]

[redacted]

The access data (QR code / link) for the survey also serve to correctly assign the questionnaires via a pseudo-anonymized key. Via this key, Frankfurt School will raffle the vouchers and, if necessary, anonymously link survey data with other key figures at the store, department or employee level (e.g. sales and/or customer frequency). Due to the described procedure, however, neither [redacted] nor [redacted] will be able to draw conclusions about survey data or key figures of individual employees.

Your **participation** in the survey **is voluntary**. You may choose not to participate. If you decide to participate in this survey, you can cancel it at any time. If you decide not to participate in this survey, or if you withdraw from participation during the survey, there will be no negative consequences.

We would like to thank you in advance for your participation and your support.

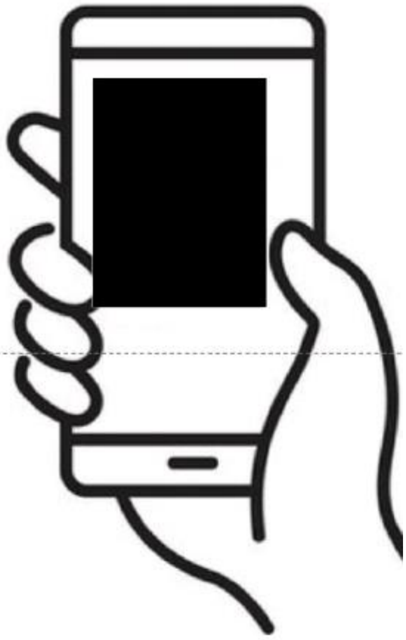
Kind regards,

[redacted]

You find the **access data to your report** again in the following:

Link: [redacted]

QR-Code: [redacted]



*Appendix B 4: The Survey*

For implementation, the survey was translated into German. To assess how the introduction of RPI affects psychological empowerment, job satisfaction, and felt job security of employees, we built on items from different scales (Fletcher & Nusbaum, 2010; Gilbert et al., 2007; MacDonald & MacIntyre, 1997; Newby & Klein, 2014; Spreitzer, 1995). We modified the wording of the items to suit our context and meet the works council's requirements, which has to agree to employee questionnaires in Germany. One requirement, e.g., was that it is evident to employees that the survey is related to the average sales report introduced.

**Employee Survey on the Average Sales Report**

In this set of statements, please indicate your level of agreement on a scale of 1 (disagree strongly) to 7 (agree strongly). Please think of the last three months when answering the questions

| | Strongly disagree | | | Neither agree nor disagree | | | Strongly agree |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1. The information I received about the average sales per transaction was interesting. | _ | _ | _ | _ | _ | _ | _ |
| 2. I would like to receive the report in the future as well. | _ | _ | _ | _ | _ | _ | _ |
| 3. I think the average sales per transaction is an important variable to evaluate my performance. | _ | _ | _ | _ | _ | _ | _ |
| 4. The information I received about the average sales per transaction helped me to evaluate my sales performance. | _ | _ | _ | _ | _ | _ | _ |
| 5. The information I received about the average sales per transaction motivated me. | _ | _ | _ | _ | _ | _ | _ |
| 6. I feel a great sense of personal satisfaction when I perform well in the average sales report. | _ | _ | _ | _ | _ | _ | _ |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7. | I feel bad and unhappy when I discover that I have performed poorly in the average sales report. | _ | _ | _ | _ | _ | _ | _ |
| 8. | I try my hardest to perform well regarding the average sales per transaction. | _ | _ | _ | _ | _ | _ | _ |
| 9. | The introduction of the average sales report has improved my overall job satisfaction. | _ | _ | _ | _ | _ | _ | _ |
| 10. | The introduction of the average sales report has improved the level of cooperation between team members. | _ | _ | _ | _ | _ | _ | _ |
| 11. | The introduction of the average sales report has improved the supportiveness of my team. | _ | _ | _ | _ | _ | _ | _ |
| 12. | Due to the average sales report my coworkers are constantly competing with one another. | _ | _ | _ | _ | _ | _ | _ |
| 13. | Due to the average sales report, I constantly feel monitored. | _ | _ | _ | _ | _ | _ | _ |
| 14. | Due to the average sales report, I feel stressed at work. | _ | _ | _ | _ | _ | _ | _ |

15. When evaluating your performance, what is most relevant to you? (Only for employees in *Median RPI* or *Distribution RPI* )

    ( ) The average sales per transaction of other employees

    ( ) The average sales per transaction I achieved in previous weeks

    ( ) Both are equally relevant

    ( ) None are relevant

16. When comparing your performance with the performance of other employees, what was most relevant to you? (Only for employees in *Distribution RPI* )

    ( ) Top 10%

    ( ) 20%

    ( ) 30%

    ( ) 40%

    ( ) 50%

    ( ) 60%

( ) 70%

( ) 80%

( ) 90%

( ) 100%

( ) None was relevant

17. When evaluating your sales performance, what is relevant to you? (Multiple answers possible)
    ( ) My average sales per transaction in the butchery department
    ( ) My average sales per transaction in selling meat
    ( ) My average sales per transaction in selling sausage
    ( ) None is relevant

18. Working in the butchery department, do you have a preference for selling meat or selling sausage?
    ( ) I prefer to sell meat
    ( ) I prefer to sell sausage
    ( ) I do not have a preference for either meat or sausage

19. Working in the butchery department, do you have specialized knowledge about the products you sell?
    ( ) I know a lot about meat
    ( ) I know a lot about sausage
    ( ) I know a lot about meat and sausage
    ( ) I do not have specialized knowledge about meat or sausage

20. Do you have any tips on how one can increase the average sales per transaction? (We anonymously collect best practices of all employees)

21. Did you have difficulties understanding the performance report presented in the app?

22. Did you have difficulties understanding the performance report presented in the app?

23. Is there any other information or performance metrics you would like to be informed about?

24. Do you have any suggestions for improvements for the app?