

A Service of



Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Kohler, Maximilian; Mahlendorf, Matthias D.; Seiter, Mischa; Vogelsang, Timo

Article — Published Version Social Comparison on Multiple Tasks: Sacrificing Overall Performance for Local Excellence?

Journal of Accounting Research

Provided in Cooperation with: John Wiley & Sons

Suggested Citation: Kohler, Maximilian; Mahlendorf, Matthias D.; Seiter, Mischa; Vogelsang, Timo (2024) : Social Comparison on Multiple Tasks: Sacrificing Overall Performance for Local Excellence?, Journal of Accounting Research, ISSN 1475-679X, Wiley, Hoboken, NJ, Vol. 62, Iss. 4, pp. 1309-1361, https://doi.org/10.1111/1475-679X.12535

This Version is available at: https://hdl.handle.net/10419/306234

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.



http://creativecommons.org/licenses/by/4.0/

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.





Journal of Accounting Research

DOI: 10.1111/1475-679X.12535 Journal of Accounting Research Vol. 62 No. 4 September 2024 Printed in U.S.A. CHICAGO BOOTH 🐺

Social Comparison on Multiple Tasks: Sacrificing Overall Performance for Local Excellence?

MAXIMILIAN KOHLER,* MATTHIAS D. MAHLENDORF,† MISCHA SEITER,* AND TIMO VOGELSANG D†

Received 30 July 2021; accepted 26 February 2024

ABSTRACT

This field experiment investigates how different levels of aggregation in relative performance information (RPI) impact employee performance in environments with multiple tasks. We randomly assign store employees of a retail chain to three groups: RPI on overall performance (control group), RPI on separate tasks, and RPI on both overall performance and separate tasks. We do not find evidence that providing separate task RPI instead of overall RPI affects performance or effort allocation. However, providing RPI on both

^{*}Institute of Business Analytics, Ulm University; [†]Department of Accounting, Frankfurt School of Finance & Management

Accepted by Haresh Sapra. We thank Haresh Sapra, an anonymous associate editor, an anonymous reviewer, Hari Ramasubramanian, Ivo Tafkov, Laurence van Lent, seminar participants at the Frankfurt School of Finance and Management, and conference participants of the *Journal of Accounting Research* Second Registered Report Conference for their helpful comments. We are very grateful for the assistance from the company. 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. This paper is the final Registered Report resulting from the Registration-based Editorial Process (REP) implemented by JAR for its special 2021 Registered Reports Conference; details of the process are available here: https://www.chicagobooth.edu/research/chookaszian/journal-of-accounting-research/registered-reports. The accepted proposal and an online appendix for this report can be downloaded at https://research.chicagobooth.edu/arc/journal-of-accounting-research/online-supplements.

^{© 2024} The Authors. *Journal of Accounting Research* published by Wiley Periodicals LLC on behalf of The Chookaszian Accounting Research Center at the University of Chicago Booth School of Business. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

overall performance and separate tasks seems to reduce performance, especially in the low-return task. This suggests that detailed RPI directs employees' attention to the smaller benefits of low-return tasks. We further find that only 30.5% of the employees accessed their performance reports, highlighting a distinction between providing RPI in the field and the laboratory. This study is based on a registered report accepted by the *Journal of Accounting Research*.

JEL codes: C93, D91, J24, M41, M50

Keywords: relative performance information; multiple tasks; social comparison; field experiment

1. Introduction

Using social comparison is a long-standing practice in organizations. In the United States, nearly one-third of organizations provide their employees with relative performance information (RPI; e.g., McGregor [2006]). Evidence from the lab (e.g., Tafkov [2013]), as well as field research on RPI (e.g., Azmat and Iriberri [2010], Blanes i Vidal and Nossol [2011], Eyring and Narayanan [2018]), show that enabling people to anonymously compare themselves with others using one performance measure can increase performance. Yet, in practice, many jobs involve a variety of different tasks and objectives (e.g., Prendergast [1999]). Crucially, digitization makes it easier to collect, analyze, and communicate such various kinds of performance information (Casas-Arce et al. [2022]). This progress increasingly enables firms to provide their employees with RPI on many relevant tasks separately. Although this increased information provision can be beneficial, it may also harm employees' attention focus and effort allocations. However, the trade-offs between providing RPI for aggregated overall performance versus providing separate RPI measures for each relevant task are a priori not clear (Hannan et al. [2019], Lu [2022]). We investigate these trade-offs by conducting a field experiment in a large German retail firm.¹

In addition to this investigation, it is essential to note that most RPI research is conducted in the laboratory (e.g., Hannan et al. [2013, 2019], Tafkov [2013]). This provides important insights into basic human behavior. However, in companies, many other things compete for the employees' attention, which is in stark difference to the laboratory and raises the question to what extent employees are interested in RPI and how it affects performance in the field. To the best of our knowledge, our project is among the first to investigate RPI that is privately provided to employees in a company setting where few incentives exist to use the RPI. Thus, this study also provides insights into how strongly social comparison incentives are in a workplace setting.

¹ In our setting, selling different products can be aggregated to overall revenue. Generally, firms often use weights on different performance dimensions to construct overall performance measures.

The experiment is based on a registered report accepted by the *Journal of Accounting Research*. Precisely, our field experiment focuses on the trade-offs of the following three possible RPI designs: (1) a single RPI measure on the aggregated overall performance of all relevant tasks (*Overall RPI*), (2) one RPI measure for each task separately (*Separate RPI*), (3) the combination of both, that is, RPI on each task separately as well as on overall performance (*Separate & Overall RPI*).

We hypothesized ex ante that providing Separate RPI can improve the informativeness of RPI, especially if the tasks have different returns to effort. With Overall RPI, employees who have to allocate a larger proportion of their effort toward tasks of low return are disadvantaged. In contrast to Overall RPI, Separate RPI avoids bundling heterogeneous tasks (Sabac and Yoo [2018]). Consequently, Separate RPI might be more relevant to achieving social distinction and could thus have a larger effect on performance (Hannan et al. [2019]). However, Separate RPI could have undesired side effects. Receiving Separate RPI, employees might focus on the task in which they are already ahead while neglecting tasks' return to effort (Hannan et al. [2013], Gill et al. [2019]). By providing Separate & Overall RPI, the firm might get the best of both worlds-increased informativeness on the one hand and emphasis on the fact that overall performance matters on the other. However, explicitly informing employees that they are outperformed overall due to an effort allocation driven by external factors Separate & Overall RPI could also demotivate (Bursztyn and Jensen [2015], Sofoklis and Megalokonomou [2021]).

We collaborate with a large German supermarket chain to execute a field experiment in the butchery departments of their stores. Employees of the butchery department have two tasks (selling meat vs. selling sausage) and work in different stores under relatively similar conditions. The average sales per transaction is 40% higher for meat than for sausage. Thus, the two categories have different returns to effort with regard to the overall average sales per transaction. Employees have practically no influence on the number of customers visiting their department and cannot choose or avoid customers. However, they can upsell customers, that is, put in additional effort, thereby considerably impacting what customers buy. For example, through competent advice and friendliness, a sales employee may persuade the customer to buy a high-priced organic chicken from a local producer instead of a discount chicken from a factory farm. Therefore, the average sales per transaction is a good indicator of sales performance.²

² The company aims for staffing where the existing staff can handle all potential customers. Accordingly, maximizing sales by quickly processing customers might be beneficial for the revenue generated per employee but not for the department's overall revenue. Instead, the company aims to ensure that employees serve customers in the best possible way. The average sales per transaction reflects how well a customer is served. Furthermore, in contrast to the total revenue generated, the average sales per transaction is comparable across employees, even though they work different hours, making it a suitable performance measure for RPI.

The company already provides employees with RPI on their average sales per transaction in the butchery department every week using a mobile app to which employees have continuous access.³

Our field experiment yields data from 410 sales employees in the butchery department. We randomly assign employees on the store level to one of three groups and manipulate the RPI information each group receives for three months. Employees in the *Overall RPI* group serve as our control group and keep receiving RPI about their overall average sales per transaction in the butchery department. Employees in the *Separate RPI* group receive only RPI on their average sales per transaction in the meat and sausage task separately. Employees in the *Overall & Separate RPI* group receive RPI on their average sales per transaction in the butchery department and RPI on their average sales per transaction in separate tasks. In all groups, only employees themselves can see their performance, that is, RPI is private.

Our study provides four key results. First, our planned analysis suggests that Separate & Overall RPI reduces performance compared to Overall RPI. The effect is statistically significant for the low-return task sausage with an approximated decrease of -2.2% in the average sales per transaction. Coefficients for the effect on overall performance are of similar size but not statistically significant in most specifications. Conducting unplanned analyses, we find that the negative effect on the low-return task is driven by employees with a high relative performance prior to the experiment. Second, our planned analyses show that in contrast to prior research in the lab, for instance, Hannan et al. [2013], we find no statistically significant evidence that providing Separate RPI instead of Overall RPI affects employees' performance or effort allocation. Thus, there is no evidence that Separate RPI causes employees to focus on local excellence at the expense of overall performance. Third, the planned survey suggests that Separate RPI is not perceived as more relevant than Overall RPI. Thus, our results suggest that making RPI objectively more informative by adding more detail on separate tasks does not necessarily imply that employees perceive it more relevant. Fourth, around one-third of the employees opened the performance reports provided at least once. Only 17.3% opened the performance reports twice or more. This questions whether employees are as interested in private RPI as some prior research assumes. It also points to the fact that, while our intention-to-treat (ITT) estimates are unbiased and important for managers when considering rolling out such an intervention, the results should be interpreted with caution as much remains to be learned about how to make RPI relevant to the majority of employees in real work settings.

Average sales per transaction is a typical metric in retail (e.g., Bullard [2016], Manthei et al. [2021]) and is considered very important by the firm.

³ The introduction of this app in the company was accompanied as part of an earlier experiment that preceded the current study (Kohler et al. [2023]).

Our field experiment makes three main contributions. First, we contribute to the literature on the effect of different RPI designs (e.g., Eyring and Narayanan [2018], Hannan et al. [2019], Knauer, Sommer, and Wöhrmann [2020], Kohler et al. [2023]) by providing evidence on the effects of the two general RPI designs Separate RPI and Separate & Overall RPI on performance and effort allocation. As benchmarking and multitasking are widely present in organizations, our findings are relevant for organizations and researchers alike. To the best of our knowledge, no prior study investigated how different aggregation levels of RPI across multiple tasks affect performance and effort allocation in the field. Our study indicates that in an environment with multiple tasks that have different returns to effort and in which the performance of individual tasks can be aggregated, Separate & Overall RPI risks to direct employees' attention to the missing benefits of low-return tasks, potentially leading to a neglect of these tasks and a reduction in performance. This also contributes to the growing literature on attention-direction in accounting. Studies have, for instance, investigated how the design of performance reports shifts attention between activities (e.g., Banker, Chang, and Pizzini [2004], Cardinaels and van Veen-Dirks [2010], Chen, Jermias, and Panggabean [2016]), how attention direction of management controls develops over time (Manthei, Sliwka, and Vogelsang [2023]), or how information overload might cause attention-allocation problems (Lu [2022]).

Second, our study is among the first to investigate the performance effects of RPI in the absence of explicit incentives within a firm using a field experiment. Although others have done substantial work in the laboratory (Kuhnen and Tymula [2012], Tafkov [2013], Hannan et al. [2013, 2019], Kramer, Maas, and Rinsum [2016]) or other field settings (Azmat and Iriberri [2010], Eyring and Narayanan [2018]), few studies were implemented in a workplace setting where employees are subject to the multiple influences of daily work life (Blanes i Vidal and Nossol [2011], Song et al. [2018]). In this regard, our study helps to calibrate the understanding of the strength of social comparison incentives in workplace settings. In our workplace setting, with little incentives and many other things competing for employees' attention, around two-thirds of the subjects chose not to access the privately provided RPI. Thus, our study highlights the importance of considering employees' motivation to access the information and ensure frictionless access to the information provided.

Finding a limited take-up of the treatment, our study also highlights an important difference between laboratory and field experiments. Although in lab experiments typically every participant is confronted with the RPI provided, treatment take-up often cannot be enforced in the field. This creates a difference between the ITT effect and the local average treatment effect (LATE). Although both are of interest, depending on the goal of the manager, policy maker, or researcher, it is important to differentiate between the two as incomplete take-up might reduce the possible effect sizes in the field after roll-out.

2. Literature Overview and Hypotheses

According to social comparison theory, people obtain utility from outperforming others (Suls and Wheeler [2000]). In the absence of objective standards, RPI allows people to learn about their performance compared to others, thus stimulating motivation to increase performance. Although evidence from the workplace is still limited, research studying the effect of private RPI in different field settings shows that private RPI can increase individuals' performance, even though performance is not linked to explicit incentives (e.g., Azmat and Iriberri [2010], Eyring and Narayanan [2018]).⁴

In settings where employees have to execute multiple tasks, companies must decide whether to provide RPI based on a single aggregate measure of the performance of all tasks (*Overall RPI*) or multiple RPIs based on the performance in each separate task (*Separate RPI*). For example, a banker selling investment products and processing loans could be evaluated based on overall profit or profit per product type; a service center employee handling customer complaints for different firms could be evaluated by the overall number of complaints or complaints per type of firm; or a programmer writing code for different applications could be evaluated based on lines of code written overall or per application type.

Providing RPI on the task level can be advantageous compared to providing only aggregated RPI. *Separate RPI* enables the firm to provide RPI without information loss and costly customization like forming peer groups (Sabac and Yoo [2018]). We argue that this is relevant because, in practice, settings with multiple tasks are often characterized by three features reducing the informativeness of aggregate RPI. (1) Different tasks can have different returns to effort. For example, selling a product in a high-priced product category can lead to more revenue with the same amount of effort as selling a product in a low-priced product category. (2) The effort allocation on separate tasks is not necessarily under the complete control of the individuals. For example, a salesperson has to serve customers' requests at a minimum. If a customer explicitly requires to buy three items of the low-return category, the employee cannot avoid selling these three items. However, individuals can usually go the

⁴Laboratory studies similarly show that RPI increases performance in the absence of explicit incentives (Kuhnen and Tymula [2012], Hannan et al. [2013], Tafkov [2013], Kramer et al. [2016]). Kramer et al. [2016] provide evidence that RPI increases performance in the absence of explicit incentives but that the performance effect does not depend on the presentation format of rank information. Hannan et al. [2013] find that RPI increases performance in multi-task environments but that this effect depends on employees' influence on task composition. Kuhnen and Tymula [2012] find that providing individuals with private RPI under a flat wage incentive scheme increases output. Tafkov (2013) shows that private RPI has a positive effect on performance and that this effect can be enhanced by providing RPI public. Furthermore, the effect of RPI is stronger under performance-based incentives than under a flat wage incentive scheme.

extra mile and provide additional effort above minimum requirements, that is, by trying to upsell. (3) Task composition (i.e., the quantity of task A vs. task B) may not be entirely homogenous among individuals within a peer group. For example, one salesperson may be "lucky" to have more customers interested in a high-priced product than another salesperson because of her position at the food counter. In other words, peer groups are not perfectly filtering out heterogeneous exogenous conditions.

Under such conditions, RPI on individual tasks separately (*Separate RPI*) is more informative about the individual's effort than a single overall performance measure (*Overall RPI*) because the overall measure is affected by heterogeneous task composition, which employees cannot fully control.

Previous research suggests that employees redirect their activities as a result of the performance measures provided (Abernethy, Bouwens, and Van Lent [2013]). Testing RPI designs with different degrees of informativeness in the laboratory, Hannan et al. [2019] find that the effect of public RPI is lower if it is less informative. It appears that a more informative RPI enables individuals to achieve greater social distinction and is thus perceived as more relevant, increasing its effect on performance. While Hannan et al. [2019] only test temporal aggregation and the level of detail, they suggest that future research could examine other RPI characteristics, such as Overall RPI versus Separate RPI. Considering the argument above, being more informative, Separate RPI can be expected to have a more positive effect on performance than Overall RPI. The increased informativeness could be especially relevant if the task composition of some employees is systematically biased. By systematically biased, we mean that by job design, some individuals in a peer group get assigned a larger proportion of the task with a higher/lower return to the effort, biasing their relative overall performance. For example, employees assigned more often to work at the meat section of a supermarket's butchery counter will have a higher overall average sales per transaction than employees who are more often assigned to the sausage section of the counter because meat is more expensive. Filtering out the effect of heterogeneous task composition, Separate RPI might avoid demotivating employees who have to allocate much effort toward low-return tasks.

However, providing *Separate RPI* also has disadvantages compared to providing *Overall RPI*. Generally, increasing the number of performance measures risks that employees excuse low performance in one task with high performance in another (Simons and Dávila [2021]). Furthermore, while private RPI on separate tasks potentially has a more positive effect on effort overall, it might cause individuals to focus on some tasks at the expense of other tasks without considering the effect on overall performance (Hannan et al. [2013]). Building on social comparison theory and self-affirmation, Hannan et al. [2013] argue that individuals strive to outperform others and can justify the negative utility of low performance in one task with

superior performance in the other task.⁵ Consequently, employees prefer excellent performance on one task and poor performance on another task to mediocre performance in separate tasks, even if it results in lower overall performance (Brickman and Bulman [1977], Hannan et al. [2013]). This is supported by Gill et al. [2019], who find that individuals increase effort most after being privately informed that they are among the best but decrease effort after being informed that they are among median performers. This mechanism could be particularly relevant in our setting because RPI is private, allowing employees to selectively disclose their performance only for tasks they perform well. Focusing effort on one task would be especially serious for the firm if that task has a lower overall return to effort.

The arguments presented above raise the question of whether providing *Separate RPI* instead of *Overall RPI* leads to additional effort on both tasks or rather an effort focused on only one task (potentially with a decrease in the other task) and how this affects performance. Given that it is not clear ex ante whether the increased motivation due to less noisy performance comparison in the *Separate RPI* condition outweighs the effect of focusing effort more on one measure, we state the following hypotheses:

- *H1a*: The provision of *Separate RPI* leads to higher overall performance than providing only *Overall RPI*.
- *H1b*: The provision of *Separate RPI* leads to lower overall performance than providing only *Overall RPI*.

A third option for the RPI designer is to provide *Separate and Overall RPI* and thus both the single *Overall RPI* as well as *Separate RPI*. Ideally, such a design could benefit from increased informativeness while emphasizing that, in the end, the tasks' return to effort and, thereby, contribution to overall performance matters, not just excelling in a single task. Even if the employee decides to focus his/her effort more on one task, the overall performance measure signals that it is more desirable to focus on the high-return tasks (i.e., the separate task that has a stronger effect on the overall measure).

However, there might also be downsides to presenting separate as well as overall measures. As explained above, *Overall RPI* might demotivate employees whose task composition is systematically biased toward low-return tasks. Such employees can outperform their peers in separate tasks but still perform poorly overall. Receiving *Separate & Overall RPI*, such employees would explicitly learn that although they performed well on separate tasks, they are outperformed by their peers overall. Knowing that one is outperformed by peers can demotivate employees and cause a reduction of effort and performance (Bursztyn and Jensen [2015], Barankay [2012], Sofoklis

⁵ Self-affirmation theory suggests that employees can compensate negative utility from poor performance in one area with superior performance in another, thereby maintaining a positive self-image (Steele [1988]).

and Megalokonomou [2021]). Explicitly knowing that this is due to a systematic bias of factors out of one's control can be expected to be even more detrimental. Furthermore, providing *Separate & Overall RPI* may lead to information overload, resulting in less information being extracted. This is especially critical in environments with multiple tasks where information is interrelated and needs to be understood cohesively (Lu [2022]).

Given the conflicting arguments, we again split our hypotheses:

- H2a: The provision of Separate & Overall RPI leads to higher overall performance than only Separate RPI.
- *H2b*: The provision of *Separate & Overall RPI* leads to lower overall performance than only *Separate RPI*.

3. Field Setting

We collect data from a large retail organization operating supermarkets in Germany.⁶ Our study focuses on employee-level performance data from 42 supermarkets in one region of the company (Southern Germany) and 410 employees. The average store of our sample is approximately 3,086 m² in size, and an average employee generates €2,416.18 in sales per week in its butchery department (table 1). In the butchery department, which is the focus of this study, goods are personally sold at fresh food counters, similar to a weekly market. The butchery department represents one long fresh counter divided into two sections (meat and sausage). Employees of the butchery department sell products from both categories (figure 1).

Figure 1 shows employees' task composition and indicates that most employees perform both tasks in similar quantities, that is, have a balanced task composition. However, a substantial share of employees exists whose task composition is either focused on the sausage task or the meat task.

Transactions are registered separately for both categories and are perceived as separate by the company and the employees. Meat is generally more expensive than sausage. Thus, on average, the average sales per transaction is 40% higher for meat than for sausage (table 1). Consequently, the return to effort for overall performance (i.e., average sales per transaction in the butchery department) is likely to be higher for meat than for sausage. However, as meat and sausage are not clear substitutes, the company wants employees to sell both categories to increase its total revenue.

Sales employees are generally responsible for stocking and presenting products as well as for customer service (i.e., selling and providing advice). Concerning customer service, employees serve customers who buy from

⁶ This setting is particularly appropriate for our research question. The firm does not use explicit incentives based on individual performance for salespeople and already provides RPI on the overall average sales per transaction in the butchery department. Moreover, the retail sector is vital for the economy. For instance, retailers employed roughly 6.3% of the U.S. workforce in 2018 (Anderson [2020]).

Desc	riptive S	Statistics			
	Ν	Mean	SD	25%	75%
Female employee (1/0)	410	0.79	0.41	1	1
Age	410	45.10	13.29	35	56
Tenure	410	10.11	9.58	3	16
Store size (m ²)	42	3,086.38	1,265.36	2,275	3,801
Length fresh food counters (m)	42	25.55	6.58	20	26
Weekly sales per employee overall (€)	404	2,416.18	1,178.27	1,669.47	3,164.49
Average sale per transaction overall (\in)	404	9.11	1.58	8.26	9.85
Weekly sales per employee meat (€)	403	1,316.04	958.28	693.60	1,750.32
Average sale per transaction meat (€)	403	10.52	1.63	9.82	11.20
Weekly sales per employee sausage (€)	401	1,116.45	711.73	582.04	1,687.21
Average sale per transaction sausage (\in)	401	7.52	1.56	7.04	8.01
Contribution Meat	404	0.53	0.28	0.39	0.65
Focus Better Task	397	0.23	0.28	0.03	0.38

TABL	E	1	
Descriptive	Ste	ıtisti	c:

This table reports descriptive statistics for the overall sample. The sales data are winsorized at the 1% and 99% percentile. The sales data cover a period of eight weeks prior to March 21, 2022. 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 during the experiment. Please note that our measure of tenure is not very precise. It measures the time of employment within the company, not the time in the specific job.



FIG. 1.—Multiple tasks in the butchery department. This figure shows the task composition of employees. Specifically, it shows the share of total transactions that employees working in the butchery department generate with the meat task over a period of eight weeks prior to March 21, 2022. The sample includes data from employees who generated sales data in at least one of the four weeks prior to the experiment (N = 410).

two product categories—meat and sausage—which have different returns to effort. Employees have practically no influence on how many customers visit their department and only limited influence on their task composition.

Employees cannot choose which task to perform as this is determined by customers' requests. Customers usually know which categories they want to buy in, determining the tasks employees have to perform. During the purchasing process, customers formulate their requests, which employees must fulfill at the minimum. However, employees can choose to put in additional effort by trying to upsell customers in the categories they buy in. That RPI might influence the upselling effort is one of the main reasons the firm was interested in our proposed field experiment.

In our setting, additional effort means giving better advice and applying sales techniques during a conversation. By putting in additional effort, employees can influence their effort level for each task and, hence, to some extent, their effort allocation. As upselling requires extra effort, not all employees engage in upselling to the same extent. Furthermore, additional effort requires customers' and employees' time, which is limited. As additional effort is limited, employees must make trade-offs when allocating additional effort to tasks.

In addition to being constrained by customer requests, employees' task composition is influenced by their position at the food counter. Employees must serve customers as they come and are not allowed to avoid specific customers. Thus, an employee's number of meat and sausage transactions strongly depends on her position at the food counter. As the company aims to ensure a good customer experience, it is required that employees are evenly distributed across the counter. Consequently, some employees are advantaged, and some are disadvantaged regarding their task composition.⁷ For the majority of employees, task composition varies little over time.⁸ Thus, it can be concluded that there is a relevant number of

⁷A natural response when providing RPI would be to consider employees' position at the food counter when forming peer groups. However, there are strong reasons preventing the company from doing so. First, different food counters in different stores are not exactly the same length, and staffing levels vary. Staffing levels vary for multiple reasons, some endogenous, for example, higher staffing level due to high revenue times and days, and exogenous reasons, for example, sick leave, parental leave, vacation, or vacancies due to a tight labor market. This makes it difficult to form 'position categories'. Furthermore, by beginning a transaction at one point of the counter and ending it at another point, employees may switch their positions during shifts. Thus, even if possible to form position categories, it would be costly for the firm to record each employee's position during the day. Second, having different peer groups based on position would mean that employees working side by side in the same store get different RPI. The company wants to avoid this because it is concerned that this would confuse employees and undermine the relevance of RPI for employees to achieve social distinction. The concern is that if being in the top 10% means something different to different employees working side by side and employees do not understand why the perceived value of being in the top 10% is low.

 $^{^{8}}$ The standard deviation of task composition was below 0.146 for 75% of the employees during the eight weeks prior the experiment.

employees whose performance is either biased upward or downward on an aggregate performance measure.

Employees do not receive any compensation or bonuses based on individual performance but a fixed salary in line with the collective wage agreement for their industry. Moreover, no financial incentives are linked to sales performance. Accordingly, an employee's relative performance does not influence his or her compensation or that of other employees. Due to an agreement with the works council, the company does not use employeelevel performance data for the performance evaluation of sales employees. Employees can advance to management positions (e.g., department manager) or become experts in their product category (e.g., meat sommelier).⁹ However, getting access to training and promotions does not greatly depend on employees' quantifiable sales performance but rather on skills like personnel management, estimating order volumes, and basic business know-how. A tournament game based on sales performance is, therefore, unlikely. Our setting can thus be regarded as free of explicit and implicit incentives concerning our performance variable, making it particularly suitable for studying the isolated effect of social comparison. The primary motivating factor resulting from RPI should be the utility employees derive from learning about their relative performance compared to others.

To summarize, we have a setting in which employees have different tasks with different returns to effort, the effort allocation per task is not under the complete control of the employees, and the task composition differs due to factors that employees cannot control. As argued above, we consider such a setting as common in practice. Whether it be a banker selling investment products, a programmer writing code for different applications, or a service center employee handling complaints for different firms, all the employees mentioned above have multiple separate tasks (i.e., the banker selling different investment products). None of them can be expected to have complete control of their effort allocation per task (i.e., a service center employee who cannot refuse to handle a specific complaint). Moreover, their task composition is likely to be driven by an external factor to some extent (i.e., orders by a supervisor or requests by a customer). Thus, we consider our research setting suitable for studying the proposed hypotheses.

The information environment before the start of the experiments is as follows. Through their department manager, employees have access to performance measures on the department level (sales, average sales per transaction, margin). Furthermore, via a mobile app, employees have access to their overall average sales per transaction of the past two weeks as well as the development of their overall average sales per transaction over the last eight weeks.¹⁰ Moreover, all employees receive RPI about their over-

 $^{^{9}}$ Each butchery department usually has a department manager, a deputy department manager, and two foremen.

¹⁰ As part of a prior experiment, a mobile app was introduced, enabling sales employees in the butchery department to view their average sales per transaction Kohler et al. [2023]. Em-

all performance in the butchery department for the last two weeks in the form of deciles.¹¹ To ensure homogeneity, three peer groups are formed based on store size. The information in the app is updated every week. Due to privacy restrictions, the reports are completely anonymous and do not allow conclusions about the individual performance data of peers, that is, RPI is private. As the company is not permitted to evaluate individual sales performance, employees' reports are not shared with department managers. Figure A1 in the appendix shows the *Overall RPI* report available for the employees before the experiment starts. Employees do not receive additional individual performance information or RPI for the separate tasks (meat and sausage) but only on overall performance prior to the experiment.

4. Experimental Design

4.1 THE TREATMENTS

We randomly assign all employees working in the butchery department on the store level to three groups that differ in the level of aggregation of information they receive. In line with prior evidence (Cardinaels [2008]), RPI is presented in the form of a figure to make it easy to understand to sales employees. Furthermore, the implication of the RPI information is explained in a text below the figure (see appendix, figure A1). The treatments are as follows:

4.1.1. Overall RPI (Butchery Department). Employees in the Overall RPI condition, which serves as our control group, only receive private RPI based on their overall average sales per transaction in the butchery department. However, they do not receive information on their average sales per transaction in separate tasks. In other words, employees in the overall group continue to receive the same information that all employees received before the start of the experiment. The Overall RPI group is our reference group, as our objective is to analyze how providing Separate RPI instead of or in addition to Overall RPI affects performance and effort allocation.

ployees in the control group of the prior experiment were informed about their average sales per transaction of the past two weeks as well the development of their average sales per transaction over the last eight weeks. Employees in the treatment groups received a report similar to the control group. In addition, they received two different designs of RPI on employees from their department from comparable stores over the last two weeks. The reports were completely anonymous and did not allow conclusions about the individual performance data of peers. At the end of the prior experiment, the *Overall RPI* design shown in the appendix was implemented for all employees (figure A1).

¹¹ Our two-week period aggregates a large number of transactions conducted over several days. Aggregating such a large number of transactions ensures the informativeness of the RPI provided during the experiment. RPI is not cumulated over multiple two-week periods because, if cumulated, employees would have little influence on their average sales per transaction after a few weeks.

4.1.2. Separate RPI (Meat and Sausage). Employees in the Separate RPI group receive private RPI informing them separately on their average sales per transaction for the two tasks: selling meat and selling sausage. However, they do not receive information on their overall average sales per transaction in the butchery department.

4.1.3. Separate & Overall RPI (Butchery Department, Meat, and Sausage). Employees in the Separate & Overall RPI group receive private RPI on their average sales per transaction in the two tasks (meat and sausage) separately, as well as RPI on the overall average sales per transaction in the butchery department.

4.2 THE IMPLEMENTATION

The experimental intervention lasted three months, from March 21, 2022 to June 19, 2022. We randomly assigned the stores to one of the three groups mentioned above. We stratified our randomization based on stores' prior average sales per transaction as well as prior treatment assignment in a different experiment. We randomized on the store level to avoid spillover and contamination effects.¹² Before the intervention, all sales employees working in the butchery department were informed that the designs of the performance reports in the app would be updated. We provided this information by distributing a staff notice via the store manager (see online appendix). To increase exposure to the intervention, a large poster explaining the app and the report was installed in the staff room. Furthermore, employees received a letter from their store manager reminding them about the app and again providing them with their access data in the middle of the experiment (see online appendix).¹³

At the end of the experiment, we invited employees of all groups to participate in a survey regarding the mobile app (see online appendix). Employees received a personal invitation letter again handed over by their store manager. One of the authors' home institutions operationalized the survey to guarantee employees' anonymity. It contained questions regarding, for instance, satisfaction, stress, and social comparison, as well as questions regarding the usage of the displayed RPI (see online appendix).¹⁴

¹² As we randomized at the store level and employees usually do not change stores, it is excluded that employees from one treatment group receive information from another treatment group.

¹³ In the company of interest, information from the administration is regularly handed over by the store manager. Thus, delivering a letter through the store manager aligns with usual procedures.

¹⁴The survey was adjusted in two areas in line with the reviewer's requests. First, where allowed by the work council, the questions were rephrased to remove the association between the survey item and the performance report. Second, additional questions were included to gather information about the described mechanisms through which our treatments affect performance.

Participation in the surveys was incentivized with the possibility of winning one of three shopping vouchers, each worth $\notin 100.^{15}$

Overall, store managers, department managers, and employees were not informed that they were part of an experiment. Moreover, given our rigorous procedure, it was also impossible for them to infer from the intervention that they were part of an experiment run by a university. This enabled us to maintain a natural environment. The firm used the word "pilot" for internal communication, which is often used in firm-level field experiments (see, e.g., Friebel et al. [2017]). By using randomization in such a natural setting to estimate the treatment effect on employees, we combine the advantages of control and realism (Floyd and List [2016], Mahlendorf and Vogelsang [2023]).

4.3 THE KEY VARIABLES

Main Effect - The Average Sales per Transaction

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

The performance measure of interest is the average sales per transaction. As explained above, it is difficult for individual employees to influence how many customers visit the butchery department and in which category they buy. However, in the respective category, employees have considerable influence on what customers buy. They can increase their average sales per transaction by 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 be the market leader with the best assortment and most competent employees. As the average sales per transaction answers the question "How much does an employee sell on average per transaction?" the company and the retail industry in general (Bullard [2016]) consider it an important indicator for sales performance. Furthermore, the company prefers average sales per transaction over alternative measures, such as total sales, because it is more comparable across employees. Employees participating in our experiment work different hours. As the number of transactions performed strongly depends on the hours worked, providing employees with RPI on total sales generated would not make sense. Moreover, the number of customers is outside the individual employees' control. For the average sales per transaction, however, it does not matter how many customers an employee serves.

Receipts are generated for customers after each transaction (separately for sausage and meat). Thus, employees receive frequent feedback on their

¹⁵ Furthermore, we placed a reminder two weeks after the initial invitation was sent to increase the response rate. Store managers received a PDF document with the employee's personnel ID and individual links to the survey. Store managers then informed employees that participation in the survey is relevant and that they can complete the survey on a computer in a separate office room during their working hours.

performance, which helps them understand how their effort translates into the output measure. Furthermore, they can compare the amount on each receipt with the RPI in the app to reflect on their relative performance immediately after each transaction.

Mechanism - Effort Allocation

In our setting, overall performance equals the total sales generated divided by the number of transactions performed. This can also be displayed as the average sales per transaction of the different tasks, weighted by an employee's task composition:

$$\begin{split} \text{Average Sale Overall}_{i,t} &= \text{Average Sale Meat}_{i,t} * \frac{\# \text{Transaction Meat}_{i,t}}{\# \text{Transaction Overall}_{i,t}} \\ &+ \text{Average Sale Sausage}_{i,t} * \frac{\# \text{Transaction Sausage}_{i,t}}{\# \text{Transaction Overall}_{i,t}}. \end{split}$$

Average Sale Overall_{*i*,*t*} is the overall average sales per transaction of employees *i* in week *t*. Average Sale Meat_{*i*,*t*} is the average sales per transaction in the meat task of employees *i* in week *t*. # Transaction Overall_{*i*,*t*} is the overall number of transactions of employees *i* in week *t*. # Transaction Meat_{*i*,*t*} is the number of meat transactions of employees *i* in week *t*.

As, in our setting, the average sales per transaction and the number of transactions per task are known, each task's contribution to overall performance can be calculated, corresponding to the share of sales generated with a task.¹⁶ It shows how many transactions and sales per transaction an employee generates with a task compared to other tasks. While the overall average sales per transaction displays how much sales employees sell on average per transaction (performance measure), the contribution indicates which task they perform (effort allocation).¹⁷

Focus Better Task_{*i*,*t*} = Contribution Better Task_{*i*,*t*} - Contribution Worse Task_{*i*,*t*}.

The variable *Focus Better Task*_{*i*,*t*} reflects how much an employee focuses on the task in which he achieves higher relative performance, that is, a higher relative performance decile. It can take values between -100% and +100%. If the overall average sales per transaction is based equally on separate tasks, *Focus Better Task*_{*i*,*t*} is 0%. If an employee focuses more on the task in which she is better, *Focus Better Task*_{*i*,*t*} is positive.

4.4 DATA COLLECTION

There are four different data sources for our analysis. First, the company's IT system generates data on sales and the number of transactions

 ${}^{16} \ Contribution \ Meat_{i,t} = \frac{Average \ Sale \ Meat_{i,t} * \frac{\# \ Transaction \ Meat_{i,t}}{\# \ Transaction \ Overall_{i,t}}}{Average \ Sale \ Overall_{i,t}} = \frac{Sales \ Meat_{i,t}}{Sales \ Overall_{i,t}}$

¹⁷As tasks have different returns to effort, contribution is only an indicator of effort allocation.

performed per sales employee.¹⁸ Second, the company's personnel records provide information on age, gender, length of service, working hours, and absenteeism. Third, the app's reporting system allows us to track how often an employee opens her average sales per transaction report. Fourth, the questionnaire obtained at the end of the intervention. We have access to the complete data set starting in August 2021.

In total, our sample consists of 410 employees working in the butchery departments of 42 different stores (table 1).¹⁹ Our sample considers all employees who worked in the butchery department for at least four weeks during the experiment. To reduce noise in the data caused by extreme outliers and thus to increase statistical power, we winsorize the data at the 1st and 99th percentile. Table 1 shows summary statistics of the data set. Table A1 in the appendix further shows that the sample characteristics do not differ over time.²⁰

5. Analyses

5.1 REGRESSION SPECIFICATION

To investigate our hypotheses, we use fixed effects difference-indifference regressions with employee and time fixed effects. The respective regression equation is:

$Y_{i,t} = \beta_0 + \beta_1 * Seperate RPI_{i,t} + \beta_2 * Seperate \mathscr{S} Overall RPI_{i,t} + X_{i,t} + \alpha_i + \delta_t + \varepsilon_{i,t}.$ (1)

 $Y_{i,t}$ denotes performance or effort allocation depending on the specification. $X_{i,t}$ captures time-variant controls.²¹ δ_t are weekly time fixed effects, α_i are individual fixed effects of the employees. The error term $\varepsilon_{i,t}$ is clustered on the unit of randomization (i.e., the store level). Seperate $RPI_{i,t}$ is a dummy variable equal to 1 for employees in the Separate RPI treatment during the experimental period, and 0 otherwise. Seperate \mathcal{E} Overall RPI_{i,t} is a dummy variable equal to 1 for employees in the Separate \mathcal{E} Overall RPI treatment during the experimental period and 0 otherwise. The reference group for our regression is the control group Overall RPI, that is, employees who only receive RPI based on their overall performance. We use data from August

¹⁸ In our company, each employee records every transaction on her individual number. If an employee sells products from the two different sections (meat and sausage), this is registered as two different transactions. Thus, the sales generated, the number of transactions performed, and the resulting average sales per transaction can be calculated separately on task level and overall.

¹⁹We expected 500 employees to participate in the experiment. However, when handing in the registered report, we did not have access to employees' work time records that are necessary to precisely estimate the number of employees included in the experiment.

²⁰ A graphical illustration of the statistical power with our sample depending on different minimum detectable effect sizes is provided in the appendix (figure A2).

²¹ Time-variant controls are the actual weekly hours worked by an employee as well as indicator variables in the pre-treatment period that capture different performance reports tested in the experiment prior to the experiment of this study (i.e., November 2021 to January 2022).

TABLE 2

	Main Treatment	Effects	
	(1) Average Sale Overall	(2) Average Sale Meat	(3) Average Sale Sausage
Separate RPI	-0.068	0.007	0.042
-	(0.151)	(0.173)	(0.115)
Separate & Overall RPI	-0.119	-0.116	-0.168^{*}
-	(0.176)	(0.196)	(0.100)
Wald test (<i>p</i> -value)	0.692	0.361	0.114
Time FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee
Number of observations	15,095	14,911	14,861
Number of employees	410	407	410
R ^e	0.515	0.445	0.463

This table reports results from a planned fixed effects regression with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 2022–June 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 *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.

2021 until the end of the experimental period in June 2022. The estimated treatment effects from the above regression give us ITT effects, which we compare to the control group and between each other using a Wald test.

5.2 MAIN RESULT

Our hypotheses focus on the overall effects of the treatments on performance. As outlined in our registered report, we thus use the above regression equation (1) with the individual overall average sales per transaction in the butchery department for employee i in week t as the dependent variable (*Average Sale Overall*_{*i*,t}) to estimate the overall effect on performance. Furthermore, we use the average sales per transaction for employee i in week t of the two different tasks (meat and sausage) as a dependent variable to estimate the effect of RPI on the performance in separate tasks. Table 2 shows the estimated average (ITT) treatment effects.

The point estimates for the treatment *Separate RPI* are neither economically nor statistically significant (table 2).²² At the same time, providing

 $^{^{22}}$ Please note, however, that the respective confidence bands are rather wide. The 95% confidence intervals for the point estimates of *Separate RPI* are [-0.373; 0.238] for the overall average sales per transaction, [-0.341; 0.356] for the average sales per transaction meat, and [-0.191; 0.276] for the average sales per transaction sausage. We, thus, do not claim that our estimates necessarily provide evidence for a null effect (William et al. [2022]).

			JJ			
	(1)	(2)	(3)	(4)	(5)	(6)
	Average Sale Overall	Decile Overall	Average Sale Meat	Decile Meat	Average Sale Sausage	Decile Sausage
Separate RPI	-0.072	-0.122	-0.001	-0.089	0.014	-0.074
	(0.110)	(0.117)	(0.123)	(0.165)	(0.053)	(0.157)
Separate & Overall RPI	-0.157	-0.243^{*}	-0.166	-0.187	-0.106^{**}	-0.445^{***}
*	(0.112)	(0.130)	(0.128)	(0.175)	(0.047)	(0.127)
Wald test (p-value)	0.368	0.243	0.106	0.413	0.046	0.026
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	No	Yes	No	Yes	No	Yes
Store FE	Yes	No	Yes	No	Yes	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Store	Employee	e Store	Employee	e Store	Employee
Number of observations	5 1,913	15,167	1,913	15,167	1,913	15,167
Number of employees	410	410	410	410	410	410
Number of stores	42	42	42	42	42	42
R^2	0.910	0.614	0.912	0.397	0.905	0.464

TABLE 3

 Further Treatment Effects

This table reports results from an unplanned fixed effects regression. In columns 1, 3, and 5, employees' average sales per transaction overall and in the meat and the sausage department on the store level is used as the dependent variable. In columns 2, 4, and 6, employees' performance decile overall and in the sausage and meat department is used as the dependent variable. The regressions account for time and employee fixed effects or time and store fixed effects. The regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 2022–June 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 *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 < 0.05, p < 0.01.

Separate & Overall RPI seems to cause a decline in performance. All coefficients for Separate & Overall RPI show negative signs for the effect on the average sales per transaction. The effect, albeit comparable large across dependent variables, is statistically significant only in the low-return task sausage. Precisely, the point estimate of Separate & Overall RPI on the average sales per transaction sausage compared to the control group Overall RPI is statistically significant with a value of -€0.17 at the employee level. Applied to the prior mean of the average sales per transaction sausage, this corresponds to a performance decline of -2.2% (table 1). Comparing the effect of Separate & Overall RPI to Separate RPI using a Wald test on the average sales per transaction in the sausage task yields a *p*-value of 0.114.

To check the robustness of our findings, we perform two unplanned analyses. First, following Athey and Imbens [2017], we use the average sales per transaction on the level of randomization, the store level *s* in week *t*, as the dependent variable. Second, we use the performance deciles of individual employees (i.e., the performance measure visible in the app) instead of absolute performance as the dependent variable. The results of the unplanned analyses are displayed in table 3. ²³

Table 3 supports our prior findings. Similar to the planned analysis on the employee level with the average sales per transaction as the dependent variable, we find no statistically significant effect for the treatment *Separate RPI* in the unplanned analyses (table 3). Furthermore, the store level analysis and the analysis of the performance deciles provide additional support for our above finding that *Separate & Overall RPI* may negatively affect performance. While all point estimates of *Separate & Overall RPI* are again negative, the point estimate is statistically significant and negative for the average sales per transaction in the sausage task at the store level -€0.11 and for the performance decile sausage -0.455 (table 3). Both point estimates differ statistically significantly between *Separate RPI* and *Separate & Overall RPI* (table 3). In addition, we find a statistically significant negative effect of *Separate & Overall RPI* on the overall performance decile of employees -0.243 (table 3).

Thus, regarding H2a/H2b, we find partial support for H2b. Providing *Separate & Overall RPI* instead of *Overall RPI* or *Separate RPI* has a statistically significant negative effect on performance in the low-return task. The effect on overall performance is also consistently negative but not statistically significant in most specifications. This suggests that *Separate & Overall RPI* negatively affects performance in the low-return task while not increasing it in the high-return task, thereby potentially harming overall performance. Regarding the effect of *Separate RPI*, we find no evidence for H1a and H1b, that is, that providing *Separate RPI* instead of *Overall RPI* has a positive or negative effect on performance.

5.3 POTENTIAL MECHANISMS

As outlined above, we find no statistically significant evidence that *Separate RPI* does affect performance compared to *Overall RPI*. At the same time, *Separate & Overall RPI* seems to have a negative effect on performance in the low-return task. The following analyses provide evidence regarding the potential underlying mechanism behind these results.

5.3.1. Effort Shift and Local Performance. Following our registered report, we use the following dependent variables in the above regression equation (1): Focus Better Task_{i,t} or Contribution Meat_{i,t}. This helps us investigate whether employees changed their effort allocation due to our treatments.²⁴

 $^{^{23}}$ We also check whether our results are biased by pre-experimental nonparallel trends by using regression equation (1) with placebo treatment variables in an unplanned analysis (table A14). The placebo treatment variables take a value of 1 for the treatment groups during the two weeks prior to the experiment. The point estimates of the placebo dummies are insignificant. Comparing pre-treatment growth rates of the control and treatment groups similarly does not show any significant differences (appendix, table A15).

²⁴ As explained in subsection 3.2, the variable *FocusBetter_{i,t}* reflects how much employee *i* focuses on the task in which she achieves higher relative performance in week *t*, that is, a higher relative performance decile. The variable *ContributionMeat_{i,t}* captures the share of overall performance generated with the meat task by employee *i* in week *t*.

Regressing Focus Better Task_{i,t} on the treatments Separate RPI and Separate & Overall RPI does not yield any significant effect (appendix, table A2). The coefficients of Separate RPI and Separate & Overall RPI are close to zero and not statistically significant. Furthermore, we do not find that our treatments had any statistically significant effect on Contribution Meat_{i,t} (appendix, table A2). Thus, no evidence exists that our treatments affect employees' effort allocation. Neither do employees shift effort to the task they perform relatively better, that is, focus on local excellence, nor do they shift effort to the high-return task due to being provided with Separate RPI or Separate & Overall RPI.

5.3.2. Effort Allocation and the Effectiveness of RPI. Consistent with our planned analysis, we then interact the treatment indicator variables with either *Focus Better Task*_{*i*,*t*} or *Contribution Meat*_{*i*,*t*} in the above regression equation (1). The coefficients for the interaction terms indicate how a change in effort allocation is associated with the performance effect of our treatments.²⁵

Both specifications do not yield statistically significant results (appendix, tables A3 and A4). Concerning H1b, this again suggests that there is no negative effort distortion effect as a result of *Separate RPI*. With regard to H2b, it suggests that the negative effect of *Separate & Overall RPI* is not driven by employees whose focus on the low-return task sausage increases during the experiment.

To capture the effect of differences in employees' effort allocation prior to the experiment, we provide an unplanned analysis.²⁶ Specifically, we interact the treatment indicator variables with *Focus Meat Prior*_i, a dummy variable indicating whether employees' effort focus on the meat task was above the median in the eight weeks prior to the experiment (appendix, table A5). Again, we do not find evidence that the negative effect of *Separate* \tilde{S} *Overall RPI* is driven by employees who focus on the low-return task sausage. With all interaction terms being negative and only statistically significant for performance in the meat task, if anything, our results suggest the opposite. *Separate* \tilde{S} *Overall RPI* seems to have a negative effect on the performance of employees with a focus on the high-return task meat.

5.3.3. Different Starting Points. Another possibility that may affect the performance effect of our treatments is the starting point, that is, the prior

²⁵ Positive coefficients for the interaction terms would indicate a decision-facilitating effect of the treatments because a change in effort focus is associated with higher overall performance. In contrast, negative interaction terms would indicate detrimental effects of employees' effort focus because the change in effort focus is associated with lower overall performance.

 $^{^{26}}$ In our registered report, we focused on changes in the effort allocation, not on differences between employees. However, to test the mechanism behind H2b, that is, whether *Separate & Overall RPI* demotivates employees with an effort focus on the low-return task sausage, we also need to analyze whether the effect differs depending on differences in the effort allocation prior to the experiment.

performance of the employees. A previous focus on the high-return task meat being associated with a negative performance effect of *Separate* \mathcal{E} *Overall RPI* could, for instance, suggest that employees who were already performing well before the experiment might relax their effort provision.

We planned to use quartile regressions to estimate the effect of different starting points. The results are shown in the appendix (figure A3). Separate \mathcal{E} Overall RPI has a statistically significant negative effect on performance in the sausage task across the performance distribution (appendix, figure A3). This provides additional insights into our main analysis. The negative effect of Separate \mathcal{E} Overall RPI on performance in the sausage task applies to the whole performance distribution. We do not find significant effects in any quartile for overall performance and performance in the meat task.

However, the quartile regression does not allow to make inferences about the individual but only about the distribution as a whole (Athey and Imbens [2017]). Therefore, we perform the following unplanned analysis: To estimate the effect of different starting points, we interact the treatment indicator variables with the dummy variable $High Performer_i$ in the above regression equation (1). $High Performer_i$ indicates whether employees' overall performance was above the median in the eight weeks prior to the experiment. The results are displayed in table 4.

All point estimates for the interaction terms of *Separate & Overall RPI* and *High Performer_i* are negative (table 4).²⁷ Similarly to the average treatment effect (table 2), the effect is most precisely measured for the average sales per transaction in the low-return task sausage with a significant decrease of -€0.41 (table 4).²⁸

In total, contrary to the mechanism suggested in H2b, *Separate & Overall RPI* does not seem to have reduced the performance of employees who are outperformed by their peers. Instead, our results suggest that the negative performance effect of *Separate & Overall RPI* can be attributed to employees with an overall high performance prior to the experiment. There is no such negative effect for high performers who receive only *Separate RPI*.

5.4 take-up of the treatment

Over three months, 30.50% (N = 125) of the employees opened their performance reports at least once (appendix, table A9).²⁹ This is similar to Eyring and Narayanan [2018], who find that 26.08% of their participants

 $^{^{27}}$ Defining *High Performer*_i based on employees' performance during the four weeks prior to the experiment or using *Top Tercile*_i and *Top Quartile*_i, dummy variables indicating if an employee's overall performance was in the top tercile or quartile during the eight weeks prior to the experiment, as interaction variables, yields similar results (appendix, tables A6–A8).

²⁸ The 95% confidence interval of the point estimate of the interaction term of *Separate* \mathcal{E} *Overall RPI* and *High Performer_i* on the overall average sales per transaction equals [-0.977; 0.294]. That of the interaction term of *Separate* \mathcal{E} *Overall RPI* and *High Performer_i* on the average sales per transaction in the meat task equals [-0.842; 0.342].

²⁹ We counted a report as "opened" if an employee accessed it at least once on a specific day and "not opened" if an employee did not open her report on a specific day.

	(1) Average Sale Overall	(2) Average Sale Meat	(3) Average Sale Sausage
Separate RPI	-0.141	0.128	-0.058
-	(0.111)	(0.173)	(0.092)
Separate & Overall RPI	0.058	0.000	0.025
-	(0.191)	(0.217)	(0.114)
Separate $RPI \times$	0.107	-0.266	0.164
High Performer $(1/0)$	(0.295)	(0.283)	(0.179)
Separate&Overall RPI ×	-0.342	-0.250	-0.413^{**}
High Performer $(1/0)$	(0.315)	(0.293)	(0.179)
Time $FE \times High$ Performer	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee
Number of observations	14,975	14,799	14,742
Number of employees	404	402	404
R^2	0.528	0.452	0.469

 TABLE 4

 Differences in the Treatment Effect Depending on Prior Performance

This table reports results from an unplanned fixed effects regression with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The treatment indicator variables are interacted with the variable *High Performer*_i, a dummy variable indicating if an employee's overall performance was above or below median during the eight weeks prior to the experiment. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 22– June 22). 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 < 0.05, p < 0.01.

accessed the provided RPI at least once during the experimental period. Only 17.3 % of the employees accessed the performance report twice or more (see column 4 in table A9 in the appendix). The limited take-up does not invalidate the identification and estimation strategy of the experiment, that is, the main ITT effects remain unbiased. However, our results should be interpreted with slight caution as a relevant share of the employees did not access the treatment.³⁰

The ITT describes the treatment effect based on treatment assignment and is often referred to as the most important estimate when it comes to program effectiveness. For example, in our setting, the ITT represents the

³⁰ The incomplete take-up might reduce the statistical power of the experiment. For this reason, we incorporate various sample sizes in our graphical illustration of the statistical power (appendix, figure A2). This shows that with a sample size of 125 and a power of 80%, the treatment effect would need to be €0.80-€0.90. Compared to a standard deviation of 1.58 for the overall average sales per transaction (table 1), this equals around half of a standard deviation or, in practical terms, one additional sausage per transaction. The power analysis presents a lower bound as nonuser/noncompliers would still increase the precision of the model as the data are helpful to estimate fixed effects more precisely. However, it should be noted that we are potentially at the margin of having enough power.

. .

TABLE 5

	(1) Average Sale	(2) Average Sale	(3) Average Sale
	Overall	Meat	Sausage
Separate RPI	-0.274	0.029	0.170
	(0.609)	(0.695)	(0.478)
Separate & Overall RPI	-0.323	-0.313	-0.463
	(0.469)	(0.522)	(0.290)
Wald test (<i>p</i> -value)	0.910	0.458	0.206
Time FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Unit of Observation	Employee	Employee	Employee
Number of observations	15,095	14,911	14,861
Number of employees	410	407	410
R^2	0.003	0.003	-0.003

This table reports results from a planned instrumental variable regression with fixed effects with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The instrument is treatment assignment. Employees of the treatment groups are considered treated when they accessed their performance reports at least once during the experiment. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 22–June 22). 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 & Overall RPI*. Robust standard errors are clustered on the store level and displayed in parentheses.

p < 0.10, p < 0.05, p < 0.05, p < 0.01.

average effect the company can expect when rolling out the app. Additionally, as planned in our registered report, we identify the usage of the app and estimate the LATE.³¹ The LATE equals the performance effect on those employees who complied with the treatment and used the app. To estimate the LATE, we implement an instrumental variable (IV) regression, which is the preferred method in the case of partial noncompliance in randomized experiments (Athey and Imbens [2017]). The instrument used is treatment assignment. In the IV regression, employees of the *Separate RPI* and *Separate & Overall RPI* conditions are considered treated when they opened the app at least once during the experiment.

Compared to the average treatment effect found in the main analysis, the estimates of the LATE are much larger (table 5), which is plausible

³¹ Comparing the observable characteristics among employees who did or did not access the treatment we document certain differences (appendix, tables A9 and A10). Such differences are also present when regressing app usage on available control variables and prior performance in an unplanned analysis (appendix, table A11). Of course, this only refers to the observable differences but may provide a first indication of further unobservable differences. These sample differences, while not invalidating our ITT estimator, would affect a simple analysis of comparing noncompliers (nonuser of the app) with compliers (user of the app) due to potential endogeneity in treatment access (i.e., self-selection) (Athey and Imbens [2017]).

because only one in three employees accessed the app at least once.³² In the *Separate RPI* condition, there is a negative point estimate for the overall average sales per transaction, while the point estimates for individual tasks are close to zero or positive. In the *Separate & Overall RPI* condition, all point estimates are again negative. The point estimate of *Separate & Overall RPI* on the low-return task sausage is -€0.46 (*p*-value = 0.118).³³

To check the robustness of the above-described mechanism of different starting points, we perform the IV regression separately for employees whose performance was above/below median during the eight weeks prior to the experiment in an unplanned analysis (table 6).³⁴ This yields a statistically significant negative point estimate for *Separate & Overall RPI* in the low-return task sausage of -€0.92 (*p*-value = 0.026), providing support for the argument that the potential negative effect of *Separate & Overall RPI* on performance in the low-return task sausage is driven by high performers.

5.5 ADDITIONAL ANALYSIS

The following provides some additional analyses that we mentioned in the registered report as well as additional analyses that came up during the review process.

5.5.1. Employee Perception. As outlined in our registered report, we investigate the effect of the treatments on further outcomes elicited from the survey, such as the relevance of RPI, satisfaction, commitment, competition, and stress. The survey questions were modified at the reviewers' request after the final round of acceptance.

A total of 66 employees participated in the survey. It appears that employees did not have difficulties understanding the performance report and that the app is the preferred communication channel for relative performance feedback (appendix, table A16, Q20 and Q22). Furthermore, the survey suggests that the majority of employees did not share their performance report with their supervisors or peers (appendix, table A16, Q14 and Q15).

Figure 2 presents the results for the survey variables considering the impact of our treatment on employees (Q1–Q12). There are two pieces of

 $^{^{32}}$ To consider potential learning effects, we provide the same analysis only considering employees treated when they accessed the app at least twice in addition to the planned analysis (appendix, table A12). This yields similar results implying that the treatment effect is similar among employees who opened the app at least twice. However, only 17.3% accessed the app at least twice again suggesting that the results should be interpreted cautiously.

³³ It seems to be acceptable to prioritize the statistical significance of the ITT estimator when comparing ITT and LATE estimators (Luca and Cole [2017]).

 $^{^{34}}$ We provide a sample split analysis because the interaction analysis performed in subsection 5.3 is difficult to implement in the instrumental variable regression. However, to align the LATE with the ITT analysis, we provide an unplanned ITT regression implementing the planned regression from table 2 separately for employees whose performance was below/above median during the eight weeks prior to the experiment (appendix, table A13).

		Low Performers			High Performers	
	(1)	(2)	(3)	(4)	(5)	(9)
	Average Sale	Average Sale	Average Sale	Average Sale	Average Sale	Average Sale
	Overall	Meat	Sausage	Overall	Meat	Sausage
Separate RPI	-0.582	0.475	-0.194	-0.158	-0.529	0.360
	(0.429)	(0.718)	(0.387)	(1.030)	(1.045)	(0.729)
Separate & Overall RPI	0.156	-0.032	0.075	-0.690	-0.617	-0.915**
	(0.606)	(0.671)	(0.356)	(0.608)	(0.623)	(0.394)
Wald test (<i>p</i> -value)	0.223	0.428	0.511	0.367	0.903	0.094
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee	Employee	Employee	Employee
Number of observations	7,563	7,434	7,500	7,412	7,365	7,242
Number of employees	205	203	205	199	199	199
Number of stores	42	42	42	42	42	42
R^{*}	0.000	0.000	0.002	0.004	0.003	-0.003
This table reports results from the sausage department as the di- below the median (1–3) and em- they opened the app at least onc	n an unplanned instrume pendent variable. The re ployees whose prior perfi e during the experiment	entral variable regression v gressions are performed s ormance was above the m and are assigned to a trea	with fixed effects with the eparately for employees v ledian (4–6). The instru- tument condition. The re	employees' average sales whose performance durin, nent is treatment assignm gressions account for time	s per transaction overall a g the eight weeks prior to lent. Employees are consider e and employee fixed effe	rd in the meat and the experiment was dered treated when cts. The regressions

	for Lou
9	E)
Ξ	(LA)
TAB	Effects
	nent.

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 & Owendt RPI*. Robust standard errors are clustered on the store level and displayed in parentheses. *p < 0.10, **p < 0.05, ***p < 0.01. compare pre-treatment observations (August 2021-mid-March 2022) with the observations during the experiment (mid-March 22-June 22). The treatment estimator thus refers to

1334

M. KOHLER, M. D. MAHLENDORF, M. SEITER, AND T. VOGELSANG



FIG. 2.—Results survey. This figure shows the average answer of employees to Q1–Q12 for different treatment groups (1 = not agree, 7 = strongly agree). 90% confidence bars are displayed. N = 66.

suggestive evidence.³⁵ First, when comparing Q1–Q4, it appears that providing *Separate RPI* in addition to or instead of *Overall RPI* does not increase its relevance to employees. In fact, *Separate RPI* is perceived as less relevant. This fits with employees stating that they care most about their overall performance, not their performance in individual tasks (appendix, table A16, Q16). Second, when comparing Q8–Q12, it seems that providing *Separate RPI* in addition to *Overall RPI* is perceived more negatively concerning workplace atmosphere.

5.5.2. Prior Performance and Experience. To estimate the interplay between prior performance and experience, we perform an unplanned analysis, splitting the sample based on employees' prior performance (*High Performer_i*) and then interacting our treatment indicator variables with employees' tenure with the company (*Tenure_i*) (appendix, table A17).³⁶ Only one point estimate, which is economically small, is statistically significant with a value of -€0.054, that is, the point estimate of *Separate*

³⁵ Unfortunately, our small survey sample size keeps us from making any statement about statistical significance.

³⁶ Please note that our measure of tenure is not very precise. It measures the time of employment within the company, not the time in the specific job.

 \mathcal{E} Overall RPI and Tenure_i on the average sales per transaction meat in the high performer sample (appendix, table A17, column 5). Thus, we conclude that tenure does not explain the negative effect of Separate \mathcal{E} Overall RPI in the low return task.

5.5.3. Gaming Behavior. To analyze whether employees tried to game the performance measure, we estimate the effect of our treatments on absolute sales and the number of transactions. In this regard, we follow the registered report using regression specification (1) with weekly absolute sales and the number of transactions as the dependent variable.³⁷ We do not find evidence that one of the treatments significantly affects one of them, implying that gaming behavior was not a concern in our setting (appendix, table A18).

6. Discussion

This study examines the trade-offs between Overall RPI, Separate RPI, and Separate & Overall RPI. Our findings from the planned analyses suggest that Separate RPI is not perceived as more relevant than Overall RPI and does not affect employees' performance or effort allocation compared to Overall RPI. Furthermore, we find that providing Separate & Overall RPI instead of Overall RPI has a significant negative effect on performance in the low-return task. The effect of Separate & Overall RPI on overall performance is also consistently negative but not statistically significant in most specifications.

We expected that *Separate & Overall RPI* could have a negative effect on performance. However, contrasting to our argument based on prior research (Bursztyn and Jensen [2015], Sofoklis and Megalokonomou [2021]), the planned analyses show that this negative effect cannot be attributed to employees who are outperformed by their peers or employees with an effort focus on the low-return task. Results from unplanned analyses suggest that high performers drive the negative performance effect. This result is consistent with high performers understanding *Separate & Overall RPI* and observing that they can perform well overall without achieving local excellence in both tasks. This suggests that *Separate & Overall RPI* risks that high performers justify a performance reduction in the low-return task with a high relative overall performance (Simons and Dávila [2021]).³⁸ If

³⁷ Although employees at the fresh food counter have relatively little influence on the number of customers, it is important to make sure that the number of customers per task does not drive changes in the average sales per transaction. For example, if an employee could avoid serving some sausage customers, that would increase the proportion of meat receipts, increasing the average sales per transaction, because meat receipts are higher on average than sausage receipts.

 $^{^{38}}$ Although our evidence supports this argument, we acknowledge that it is also conceivable that high performers might just neglect the low-return task because *Separate & Overall RPI* causes information overload and distracts their attention.

employees only reduce performance in one task while not increasing it in another, this has negative consequences for the firm. Although not always statistically significant at conventional levels, we find negative coefficients of a similar size for the overall performance effect of *Separate & Overall PRI* across all specifications.

In contrast to Hannan et al. [2013] and Gill et al. [2019], the planned analysis does not provide evidence that *Separate RPI* risks employees to strive for local excellence at the expense of overall performance. This might be due to differences in the experimental setting. In Hannan et al. [2013], participants have free choice over how much of their available time they spend on one task. They can even choose to ignore one of the two tasks. In Gill et al. [2019], the time spent on each task is fixed, but participants can choose whether to provide effort or not during the respective time frame. In our field setting, this choice is limited. Employees can exert more effort in a specific transaction by upselling customers. However, the time budget of their customers is not endless, and employees have to serve the next customer at some point. Furthermore, they have to serve customers' requests at a minimum. Thus, in the field, potential effort distortion effects of *Separate RPI* toward local excellence might be less severe because the effort allocation of employees is more constrained due to factors they cannot control.

With regard to Şabac and Yoo [2018] and Hannan et al. [2019], our planned analyses also highlight that higher informativeness does not necessarily equate to greater perceived relevance for social distinction. This may be due to differences in the information provided. Although Hannan et al. [2019] modify the informativeness of *Separate RPI*, for example, by varying between rank and absolute performance, our experiment tests different aggregation levels of RPI, similar to the theoretical model by Şabac and Yoo [2018]. If employees do not care about RPI in individual tasks but only about *Overall RPI*, even though it might be biased, it is plausible that *Separate RPI* is not perceived as more relevant. Further research might find it valuable to examine why employees prefer *Overall RPI* even though *Separate RPI* is objectively a fairer indicator of their performance.

Taken together, our results suggest that firms should be careful about RPI's level of detail when multiple tasks can be aggregated on an overall performance measure. Our study shows that combining *Separate RPI* and *Overall RPI* risks to direct employees' attention to the missing benefit of the low-return task. This conclusion is further supported by the suggestive evidence of our survey, which indicates that separate measures seem to be diametral to the workplace atmosphere.

7. Conclusion

Many jobs in modern economies require multiple tasks to execute. In simpler jobs with only one important task, it has been shown that RPI can be performance-enhancing. However, designing RPI for multiple tasks is difficult and, if designed wrongly, may even harm performance. Our field

experiment sheds light on the design of RPI in settings with multiple tasks. Specifically, it presents results about how detailed the level of information provided should be—on the single tasks (*Separate RPI*), on the aggregate performance (*Overall RPI*), or both (*Separate & Overall RPI*).

Our study has four main results. First, our planned analyses show that Separate & Overall RPI results in lower performance in the low-return task than providing only Separate RPI or Overall RPI. The effect on overall performance is also consistently negative but not statistically significant in most specifications. Unplanned analyses reveal that the negative effect of the lowreturn task is driven by employees with a high relative overall performance prior to the experiment. Receiving Separate & Overall RPI, such employees might observe that they can perform well overall without achieving local excellence in both tasks, consequently reducing effort in the low-return task. Second, the planned analyses suggest that providing Separate RPI instead of Overall RPI does not affect employees' performance or effort allocation. Thus, contrary to laboratory studies, we do not find that Separate RPI causes employees to focus on local excellence at the expense of overall performance. Third, although objectively more informative, the planned survey indicates that providing Separate RPI instead of or in addition to Overall RPI does not increase its perceived relevance among employees. Fourth, we discover that only 30.5 % of the sample accessed the performance reports provided at least once and 17.3 % twice or more.

The incomplete take-up of the treatment replicates results from prior field research (Eyring and Narayanan [2018], Kohler et al. [2023]) but is still remarkable considering that the RPI is communicated through a convenient digital performance report and employees received a reminder in the middle of the experiment. Although the ITT estimate of our treatments remains unbiased, the results should be interpreted with caution as much remains to be learned about what makes RPI relevant to employees in real work settings.³⁹ Currently, field studies where "nonuse" is possible regularly cannot measure usage because RPI is provided analogue in the form of a letter or poster (e.g., Blanes i Vidal and Nossol [2011], Blader, Gartenberg, and Prat [2020]). Our study highlights that future research testing encouragement designs in the field should pay attention to the actual take-up rate of the treatment, as employees' access to the information is a potentially important determinant of the performance metric effect.

The incomplete take-up is also an important point that distinguishes field experiments from laboratory experiments. Although in the lab the researcher has high control over what subjects see, people in the field may

³⁹ Potential reasons for the low usage may be that the RPI presented is private and that there are no incentives based on relative performance (Tafkov [2013]). Although this was consciously to focus on the unbiased effect of social comparison, it may have limited the potential effect of our intervention. Another potential explanation could be that employees do not feel affiliated to peers from other stores that may reduce the relevance of the RPI presented (Knippenberg [2000], Mahlendorf et al. [2014]).

choose not to look at their performance reports. Thus, while providing important insight into the behavioral mechanism, laboratory experiments might overestimate the actual effect of RPI in the field. A difference that only occasionally receives focus in the literature. Future research might integrate employees' motivation to use RPI in its model and provide the LATE as well as the ITT effect. Generally, policy makers and firms might be more interested in the ITT as this gives them the average effect of their intervention. In contrast, researchers and designers of such interventions might be more interested in the LATE as this displays the full potential of the intervention in case of complete take-up. It is, however, important to be transparent about both effects.

Finally, a limitation of the current study is that we do not have a no RPI condition and thus cannot state the baseline effect of *Overall RPI*. As recent evidence on the performance effect of private RPI in company settings is inconclusive (e.g., Blanes i Vidal and Nossol [2011], Ashraf [2022]), future research might find it valuable to further examine the effects of *Separate RPI* and *Overall RPI* in the presence of a no RPI condition.

Initially, we raised the question: Sacrificing overall performance for local excellence? Our study does not provide evidence for this potential caveat when implementing RPI in a setting with multiple tasks. In contrast, our results suggest that a more detailed RPI may risk directing employees' attention to the missing benefits of low-return tasks, potentially leading to a neglect of local performance in these tasks. Therefore, on a more general level, our study shows that in a setting where employees face multiple tasks and performance in separate tasks can be aggregated, more detail does not necessarily provide an advantage. *Separate RPI* does not provide additional value over *Overall RPI*. Providing *Separate RPI* in addition to *Overall RPI* can reduce performance.

Thus, while modern technology enables more and more opportunities to provide data to facilitate decision-making or to increase motivation, managers have to think even harder about which data to focus on. More is not always better.

APPENDIX

Time		Ν	Mean	SD	25%	75%
Week 1-4	Female employee $(1/0)$	410	0.790	0.408	1	1
	Age	410	45.098	13.287	35	56
	Tenure	410	10.107	9.582	3	16
Week 5–8	Female employee $(1/0)$	406	0.788	0.409	1	1
	Age	406	45.106	13.289	35	56
	Tenure	406	10.076	9.586	3	16
Week 9–13	Female employee $(1/0)$	393	0.786	0.410	1	1
	Age	393	44.929	13.440	34	56
	Tenure	393	10.102	9.649	3	16

ТА	BLE	A 1	
Descriptizie	Statisti	~ C (7)	r Time

This table reports descriptive statistics for the overall sample for different time frames during the experiment. The data cover a period of 13 from mid-March 22 to June 2022. Observations are excluded when an employee was absent in the respective period, when an employee did not generate sales data in the respective period, or when an employee had less than four weeks of sales data in which he worked during the experiment. Please note that our measure of tenure is not very precise. It measures the time of employment within the company, not the time in the specific job.

	(1)	(2)
	Focus Better Task	Contribution Meat
Separate RPI	-0.027	-0.006
*	(0.027)	(0.016)
Separate & Overall RPI	-0.012	0.012
•	(0.029)	(0.010)
Wald test (<i>p</i> -value)	0.615	0.259
Time FE	Yes	Yes
Individual FE	Yes	Yes
Controls	Yes	Yes
Unit of observation	Employee	Employee
Number of observations	12,475	15,095
Number of employees	407	410
R^2	0.224	0.744

TABLE A2
Treatment Effect on Effort Allocation

This table reports results from a planned fixed effects regression with the employees' *Focus Better Task*_{*i*,*t*} and *Contribution Meat*_{*i*,*t*} as the dependent variable. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 22–June 22). 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 *Separate RPI and Separate & Overall RPI*. Robust standard errors are clustered on the store level and displayed in parentheses.

	JJ J I J		
	(1) Average Sale Overall	(3) Average Sale Meat	(3) Average Sale Sausage
Separate RPI	-0.097	-0.031	-0.073
-	(0.0866)	(0.133)	(0.136)
Separate & Overall RPI	-0.177	-0.152	-0.204^{*}
-	(0.106)	(0.143)	(0.119)
Separate $RPI \times$	-0.203	-0.072	0.310
Focus Better Task	(0.229)	(0.329)	(0.219)
Separate & Overall RPI ×	0.039	0.151	0.196
Focus Better Task	(0.242)	(0.352)	(0.256)
Time FE ×	Yes	Yes	Yes
Focus Better Task			
Controls	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee
Number of observations	12,475	12,475	12,475
Number of employees	407	407	407
R^2	0.674	0.487	0.478

TABLE A3

 Treatment Effect of Employees' Focus Better Task

This table reports results from a planned fixed effects regression with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The treatment indicator variables are interacted with the variable *Focus Better Task*_{*i*,*t*}. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021–mid March 2022) with the observations during the experiment (mid-March 22–June 22). The treatment and interaction 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.

1342	M. KOHLER, M. I	. MAHLENDORF, M.	SEITER, AND T.	VOGELSANG
------	-----------------	------------------	----------------	-----------

	(1) Average Sale Overall	(2) Average Sale Meat	(3) Average Sale Sausage
Separate RPI	-0.124	0.295	0.018
	(0.262)	(0.363)	(0.215)
Separate & Overall RPI	-0.188	0.030	0.276
	(0.508)	(0.545)	(0.265)
Separate RPI \times	0.184	-0.450	0.060
Contribution Meat	(0.612)	(0.726)	(0.436)
Separate & Overall RPI ×	0.015	-0.335	-0.764
Contribution Meat	(1.090)	(1.058)	(0.512)
Time $FE \times Contribution Meat$	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee
Number of observations	15,095	14,911	14,861
Number of employees	410	407	410
R^2	0.612	0.502	0.478

TABLE A	4
Treatment Effect of Employees'	Contribution Meat

This table reports results from a planned fixed effects regression with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The treatment indicator variables are interacted with the variable *Contribution Meal*_{i,t}. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 22–June 22). The treatment and interaction 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.

	(1) Average Sale Overall	(2) Average Sale Meat	(3) Average Sale Sausage
Separate RPI	-0.021	0.212	-0.036
	(0.104)	(0.161)	(0.076)
Separate & Overall RPI	0.142	0.161	-0.050
	(0.168)	(0.182)	(0.086)
Separate RPI \times	-0.076	-0.434	0.170
Focus Meat $(1/0)$	(0.273)	(0.283)	(0.193)
Separate & Overall ×	-0.498	-0.599^{*}	-0.226
Focus Meat (1/0)	(0.301)	(0.298)	(0.231)
Time FE \times Focus Meat	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee
Number of observations	14,975	14,799	14,742
Number of employees	404	402	404
R^2	0.529	0.452	0.470

 TABLE A5
 Differences in the Treatment Effect Depending on Prior Effort Allocation

This table reports results from an unplanned fixed effects regression with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The treatment indicator variables are interacted with the variable *Focus Meat_i*, a dummy variable indicating whether employees focused on the meat or sausage task during the eight weeks prior to the experiment. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 22–June 22). The treatment and interaction 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.

Differences in the Treatment Effect Depending on Prior Performance				
	(1) Average Sale Overall	(2) Average Sale Meat	(3) Average Sale Sausage	
Separate RPI	-0.164	0.018	-0.079	
	(0.120)	(0.192)	(0.087)	
Separate & Overall RPI	0.072	0.045	0.010	
	(0.193)	(0.217)	(0.114)	
Separate RPI \times	0.174	-0.042	0.204	
High Performer $(1/0)$	(0.299)	(0.332)	(0.188)	
Separate&Overall RPI ×	-0.363	-0.340	-0.366^{*}	
High Performer $(1/0)$	(0.318)	(0.350)	(0.187)	
Time FE \times High Performer	Yes	Yes	Yes	
Individual FE	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	
Unit of observation	Employee	Employee	Employee	
Number of observations	14,887	14,729	14,655	
Number of employees	400	400	400	
R^2	0.524	0.451	0.469	

	TABLE A6	
man and in	the Treatment Effect Debending on Prior	D

This table reports results from an unplanned fixed effects regression with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The treatment indicator variables are interacted with the variable *High Performer*_i, a dummy variable indicating if an employee's overall performance was above or below median during the four weeks prior the experiment. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 22–June 22). 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.

	(1) Average Sale Overall	(2) Average Sale Meat	(3) Average Sale Sausage
Separate RPI	-0.107	0.052	-0.044
	(0.110)	(0.150)	(0.0721)
Separate & Overall RPI	-0.015	-0.066	-0.015
	(0.161)	(0.192)	(0.093)
Separate RPI \times	0.007	-0.184	0.140
Top Tercile $(1/0)$	(0.424)	(0.392)	(0.285)
Separate&Overall RPI ×	-0.306	-0.183	-0.525^{*}
Top Tercile $(1/0)$	(0.425)	(0.432)	(0.266)
Time FE × Top Performer	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee
Number of observations	14,975	14,799	14,742
Number of employees	404	402	404
R^2	0.531	0.455	0.472

TABLE A7 Differences in the Treatment Effect Depending on Prior Performance (Top Tercile)

This table reports results from an unplanned fixed effects regression with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The treatment indicator variables are interacted with the variable Top Tercile, a dummy variable indicating if an employee's overall performance was in the top tercile during the eight weeks prior the experiment. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021-mid-March 2022) with the observations during the experiment (mid-March 22-June 22). 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.

Differences in the Treatment Effect Depending on Prior Performance (Top Quartile)				
	(1) Average Sale Overall	(2) Average Sale Meat	(3) Average Sale Sausage	
Separate RPI	-0.125 (0.117)	0.052 (0.155)	-0.054 (0.067)	
Separate & Overall RPI	-0.041 (0.157)	-0.050 (0.186)	-0.033 (0.086)	
Separate RPI \times	0.125	-0.227	0.259	
Top Quartile $(1/0)$	(0.484)	(0.436)	(0.324)	
Separate&Overall RPI ×	-0.313	-0.316	-0.644^{*}	
Top Quartile $(1/0)$	(0.494)	(0.484)	(0.344)	
Time $FE \times Top$ Performer	Yes	Yes	Yes	
Individual FE	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	
Unit of observation	Employee	Employee	Employee	
Number of observations	14,975	14,799	14,742	
Number of employees	404	402	404	
R^2	0.533	0.455	0.473	

 TABLE A8

 nt Effect Debending on Prior Performance (Tob Quartile)

This table reports results from an unplanned fixed effects regression with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The treatment indicator variables are interacted with the variable $Top Quartile_i$, a dummy variable indicating if an employee's overall performance was in the top quartile during the eight weeks prior the experiment. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observa-tions (August 2021–mid-March 2022) with the observations during the experiment (mid-March 22–June 22). 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.

Oser versus ivonuser of the App				
	(1)	(2)	(3) User (Accessed App at Least	(4) User (Accessed App Twice or
	ALL	Nonuser	Once)	More)
Female employee (1/0)	0.790	0.804	0.760	0.746
. ,	(0.408)	(0.398)	(0.429)	(0.438)
Age	45.100	46.407	42.112***	45.423
-	(13.276)	(13.060)	(13.370)	(12.377)
Tenure	10.107	10.596	8.992	10.310
	(9.582)	(9.647)	(9.376)	(10.357)
High Performer $(0/1)$	0.493	0.486	0.508	0.563
	(0.501)	(0.501)	(0.502)	(0.499)
Average Sale Overall	8.962	8.946	8.997	9.037
	(1.658)	(1.796)	(1.300)	(1.333)
Average Sale Meat	10.446	10.418	10.510	10.488
_	(1.692)	(1.865)	(1.216)	(1.273)
Average Sale Sausage	7.420	7.441	7.440	7.463
	(1.478)	(1.648)	(0.994)	(1.117)
Contribution Meat	0.525	0.524	0.529	0.547
	(0.224)	(0.232)	(0.208)	(0.206)
Focus Better Task	0.232	0.235	0.227	0.225
	(0.279)	(0.290)	(0.252)	(0.257)
Number of employees	410	285	125	71
Number of employees %	1	0.695	0.305	0.173

TABLE A9 User Versus Nonuser of the App

This table reports means of the respective variables for the overall sample with standard deviations in where the set of the operation of the o

Nonuser Versus User Across Treatments				
	(1)	(2)	(3)	(4)
	Nonuser	User Control	User Separate	and Overall
Female employee (1/0)	0.804	0.760	0.875	0.674^{*}
	(0.398)	(0.431)	(0.336)	(0.474)
Age	46.407	43.060^{*}	43.281	40.140***
0	(13.060)	(12.476)	(14.328)	(13.730)
Tenure	10.596	8.200	9.062	9.860
	(9.647)	(9.269)	(9.329)	(9.672)
High Performer $(0/1)$	0.486	0.520	0.500	0.500
0	(0.501)	(0.505)	(0.508)	(0.506)
Average Sale Overall	8.946	9.098	8.569	9.199
0	(1.796)	(1.387)	(1.342)	(1.104)
Average Sale Meat	10.418	10.602	10.084	10.721
0	(1.865)	(1.165)	(1.369)	(1.101)
Average Sale Sausage	7.411	7.316	7.431	7.592
0	(1.648)	(0.770)	(1.120)	(1.123)
Contribution Meat	0.524	0.530	0.500	0.549
	(0.232)	(0.242)	(0.210)	(0.160)
Focus Better	0.235	0.271	0.199	0.195
	(0.290)	(0.273)	(0.181)	(0.268)
Number of employees ^a	285	50	32	43

TABLE A10

This table reports means of the respective variables for the overall sample with standard deviations in parentheses. *High Performer*_i is a dummy variable indicating if an employee's overall performance was above or below the median during the eight weeks prior to the experiment. The sales data are winsorized at the 1% and 99% percentile. The data cover a period of eight weeks prior March 21, 2022. *F*Test compares user to nonuser.

^aOn measures describing prior performance, the number of employees is slightly lower as not every employee included in the experiment worked prior to the experiment.

 $p^* < 0.10, p^* < 0.05, p^* < 0.01.$

Drivers of App Usage			
	(1) Active Once	(2) Active Twice	(3) Active N
High Performer (1/0)	0.032	0.062	0.377
-	(0.055)	(0.044)	(0.227)
Contribution Meat	-0.064	-0.045	-0.744
	(0.143)	(0.105)	(0.637)
Weekly Hours Worked	0.001	0.001	0.004
	(0.003)	(0.002)	(0.012)
Gender	0.001	0.041	0.299
	(0.072)	(0.066)	(0.385)
Age	-0.005^{**}	0.000	0.007
-	(0.002)	(0.001)	(0.006)
Tenure	-0.001	0.000	-0.011
	(0.003)	(0.002)	(0.009)
Counter length (m)	0.004	0.002	0.004
0	(0.003)	(0.002)	(0.010)
Store size (m ²)	-0.000	-0.000^{*}	-0.000^{**}
	(0.000)	(0.000)	(0.000)
Number of observations	404	404	404
R^2	0.028	0.014	0.019

TABLE A11 Drivers of App Usao

This 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. High Performer_i is a dummy variable indicating if an employee's overall performance was above or below median during the eight weeks prior to the experiment. The data are 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.

Local Average Treatment	Effects (LATE): Employee	es Who Accessed the App a	t Least Twice
	(1) Average Sale Overall	(2) Average Sale Meat	(3) Average Sale Sausage
Separate RPI	-0.398	0.045	0.251
	(0.889)	(1.015)	(0.710)
Separate & Overall RPI	-0.805	-0.785	-1.191
	(1.161)	(1.309)	(0.764)
Wald test (<i>p</i> -value)	0.619	0.343	0.122
Time FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee
Number of observations	15,095	14,911	14,861
Number of employees	410	407	410
R ^e	0.003	0.003	-0.008

TABLE A12

This table reports results from an unplanned instrumental variable regression with fixed effects with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The instrument is treatment assignment. Employees are considered treated when they opened the app at least twice during the experiment and are signed to a treatment condition. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021-mid-March 2022) with the observations during the experiment (mid-March 22-June 22). 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.

 $p^{*} p < 0.10, p^{*} p < 0.05, p^{*} p < 0.01.$

		Low Performers			High Performers	
	(1)	(2)	(3)	(4)	(5)	(9)
	Average Sale	Average Sale	Average Sale	Average Sale	Average Sale	Average Sale
	Overall	Meat	Sausage	Overall	Meat	Sausage
Separate RPI	-0.143	0.118	-0.047	-0.039	-0.133	0.093
1	(0.108)	(0.171)	(0.093)	(0.261)	(0.259)	(0.170)
Separate & Overall RPI	0.048	-0.009	0.023	-0.290	-0.258	-0.387^{**}
	(0.193)	(0.219)	(0.112)	(0.258)	(0.260)	(0.159)
Wald test (p -value)	0.303	0.472	0.529	0.086	0.507	0.022
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee	Employee	Employee	Employee
Number of observations	7,563	7,434	7,500	7,412	7,365	7,242
Number of employees	205	203	205	199	199	199
Number of stores	42	42	42	42	42	42
R^{*}	0.318	0.337	0.308	0.468	0.450	0.485

TABLE A13 Main Treatment Effects for Low-/High Performers

(August 2021–mid-March 2022) with the observations during the experiment (mid-March 2022–June 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 *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. employees whose prior performance was above median (4-6). The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations

	Pre-Treatment Place	ebo Test	
	(1) Average Sale Overall	(2) Average Sale Meat	(3) Average Sale Sausage
Separate RPI	-0.097	-0.381	-0.154
	(0.191)	(0.261)	(0.129)
Separate & Overall RPI	0.126	0.664	0.162
	(0.360)	(0.479)	(0.227)
Time FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee
Number of observations	10,750	10,642	10,595
Number of employees	405	404	405
R^2	0.543	0.452	0.508

TABLE A14

This table reports results from an unplanned fixed effects regression with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations beginning in August 2021 with the two weeks prior to the experiment (i.e., the experiment started mid-March 2022). The treatment estimators thus refer to the difference-in-difference estimator and simulates the treatment effect during the two weeks prior to the experiment. 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 during the experimental period (mid-March 2022–June 2022). Robust standard errors are clustered on the store level and displayed in parentheses.

*p < 0.10, **p < 0.05, ***p < 0.01.

TABLE A15

Comparing Pre-Treatment Growth Rates

	1 8		
	(1) Average Sale Overall Growth	(2) Average Sale Meat Growth	(3) Average Sale Sausage Growth
Separate RPI	-0.002	0.010	-0.009
-	(0.014)	(0.014)	(0.016)
Separate & Overall RPI	-0.007	-0.009	-0.022
•	(0.007)	(0.014)	(0.016)
Time FE	No	No	No
Individual FE	No	No	No
Controls	No	No	No
Unit of observation	Employee	Employee	Employee
Number of observations	398	396	395
Number of employees	398	396	395
R^2	0.001	0.004	0.006

This table reports results from an unplanned OLS regression with employees' growth rate of the average sales per transaction prior to the experiment as the dependent variable, that is, comparing the mean average sales per transaction during mid-February to mid-March with the mean average sales per transaction during mid-February. The regression is performed with collapsed data on the employee level. The treatment estimators thus refer to the OLS estimator and indicate whether the pre-treatment growth rate differs between the control and the treatment groups. 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 during the experimental period (mid-March 2022–June 2022). Robust standard errors are clustered on the store level and displayed in parentheses.

Q14: Sharing with Supervisor	Ν	%
Only when performed well	0	0
Only when performed poorly	0	0
Yes, independent of performance	16	0.25
No, independent of performance	48	0.75
Q15: Sharing with Colleagues	Ν	%
Only when performed well	3	0.05
Only when performed poorly	1	0.02
Yes, independent of performance	13	0.20
	40	0.74
Q16: Relevance Performance in Task	N	%
Important overall (Butchery)	31	0.52
Important separate (Meat \lor Sausage)	5	0.08
Important separate and overall (Butchery \land (Meat \lor Sausage)) Important nothing	6 17	0.01
O17: Task Preference	N	%
Meat	90	0.46
Sausage	11	0.40
No preference	23	0.37
Q18: Task Knowledge	N	%
Meat	6	0.10
Sausage	6	0.10
Both	47	0.77
No knowledge	2	0.03
Q19: Useful for Increasing the Average Sales per Transaction	Ν	%
Information from the Intranet	11	0.11
Tips from colleagues and supervisors	26	0.27
The individual receipts after each transaction	13	0.14
Knowing when I had potential for improvement compared to others	6	0.06
The motivation to see, whether I am among the Top 10%, 20%,, 100%	19	0.20
10% 20% = 100%	Э	0.05
Other	7	0.07
None of the information was relevant	11	0.11
Q20: Preferred Medium for Performance Feedback	Ν	%
Арр	39	0.49
Letter	17	0.21
SMS	5	0.06
Feedback from supervisor	8	0.10
Other	3	0.04
Do not want in future	8	0.10
Q22: Problems Understanding the Performance Report	Ν	%
No	33	0.92
Yes	3	0.08
Other	0	0.00

TABLE A16 Further Survey Results

This table shows the results of the multiple choice survey questions Q14–Q20 and Q22.

	Differences in 1	the Treatment Effect Def	bending on Prior Perfor	mance and Tenure		
	(1)	(2) Low Performer	(3)	(4)	(5) High Performer	(9)
	Average Sale Overall	Average Sale Meat	Average Sale Sausage	Average Sale Overall	Average Sale Meat	Average Sale Sausage
Separate RPI	-0.124	0.193	0.026	0.105	0.171	0.026
Separate & Overall RPI	-0.171	-0.045	-0.022	0.053	0.268	-0.353
Sebarate RPI × Temure	(0.160) -0.001	(0.221) -0.007	(0.129) -0.007	(0.202) -0.014	(0.282) -0.030	(0.270)
<i>II</i>	(0.00)	(0.017)	(0.00)	(0.025)	(0.020)	(0.020)
Separate&Overall RPI × Tenure	0.022	0.003	0.004	-0.035	-0.054^{***}	-0.002
	(0.016)	(0.020)	(0.010)	(0.022)	(0.019)	(0.018)
Time $FE \times Tenure$	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee	Employee	Employee	Employee
Number of employees	205	203	205	199	199	199
Number of observations	7,563	7,434	7,500	7,412	7,365	7,242
R^{\prime}	0.326	0.342	0.313	0.473	0.458	0.491
This table reports results from an u as the dependent variable. The sample to the experiment. The treatment indic employee fixed effects. The regressions 4 The treatment estimators thus refer to 4 thad less than four weeks of sales data in had less than four weeks of sales data in $*p < 0.10, **p < 0.01, ***p < 0.01$.	nplanned fixed effects r is split into low and hig ator variables are intera compare pre-treatment of he difference-in-differen which he worked. Robus	egression with the emp h performers, defined a cted with the variable <i>T</i> observations (Augus 20) ce estimator. Observatic st standard errors are ch	loyces' average sales pe is employees' performa: <i>inure</i> , specifying emplo 21-mid-March 2022) will an are excluded when a ustered on the store leve	r transaction overall an ree being below or abo yees' tenure in the con the observations duri n employee was absent I and displayed in pare'	id in the meat and the sa- ve the median during the apany. The regressions acc ing the experiment (mid-M in the respective week or v atheses.	usage department eight weeks prior ount for time and arch 22-June 22), when an employee

TABLE A17

		JJ		J		
	(1) Sales Overall	(2) Sales Meat	(3) Sales Sausage	(4) #Transactions Overall	(5) #Transactions Meat	(6) #Transactions Sausage
Separate RPI	8.653	-16.500	25.980	0.322	-1.847	2.265
	(84.270)	(67.330)	(54.290)	(7.917)	(4.770)	(6.722)
Separate & Overall RPI	-68.410	-42.950	-23.730	-2.696	-1.297	-1.238
	(66.990)	(54.750)	(41.490)	(7.155)	(3.601)	(5.216)
Wald test (p-value)	0.349	0.675	0.356	0.718	0.909	0.598
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Unit of observation	Employee	Employee	Employee	Employee	Employee	Employee
Observations	15,095	14,911	14,861	15,095	14,911	14,861
Number of employees	410	407	410	410	407	410
R^2	0.686	0.691	0.730	0.701	0.704	0.743

 TABLE A18
 Treatment Effect on Sales and the Number of Transactions

This table reports results from a planned fixed effects regression with the employees' weekly sales and number of transactions overall and in the meat and the sausage department as the dependent variable. The regressions account for time and employee fixed effects. The regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 202–June 22). 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 *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.



Overall RPI (Control Group)

FIG. A1.—The app and the treatments.



FIG. A2.—Power analysis. The figure reports the statistical power achieved depending on different minimum detectable effect sizes (MDE). The power analysis was executed with the Stata package pcpanel (Burlig, Louis, and Woerman [2020]), which simulates a randomized experiment with panel data using an existing data set. The data cover 24 weeks of data prior to the experiment. The 24 weeks are divided into 12 pre- and 12 post-experimental periods. Week and individual fixed effects are included. Standard errors are clustered on the store level (the unit of randomization).



FIG. A3. — Quartile regressions (Butchery, Sausage, Meat). This figure displays the results from planned quartile regressions at the 25th, 50th, and 75th quantile, with the employees' average sales per transaction overall and in the meat and the sausage department as the dependent variable. The regressions account for employee fixed effects. The quartile regressions compare pre-treatment observations (August 2021–mid-March 2022) with the observations during the experiment (mid-March 22–June 22). The treatment estimators thus refer to the difference-in-difference estimator. All regressions control for employees' work hours, report designs prior to the experiment, and time effects. 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 bootstrapped and clustered on the store level. 90% confidence bars are displayed.

REFERENCES

- ABERNETHY, M. A.; J. BOUWENS; and L. VAN LENT. "The Role of Performance Measures in the Intertemporal Decisions of Business Unit Managers." *Contemporary Accounting Research* 30 (2013): 925–61. https://doi.org/10.1111/j.1911-3846.2012.01178.x
- ANDERSON, D. A. "Retail Jobs Among The Most Common Occupations." United States Census Bureau. 2020. Available at https://www.census.gov/library/stories/2020/09/profile-of-theretail-workforce.html
- ASHRAF, A. "Performance Ranks, Conformity, and Cooperation: Evidence from a Sweater Factory." CESifo Working Paper No. 9591, Ludwig Maximilian University of Munich (LMU). 2022. https://doi.org/10.2139/ssrn.4042435
- ATHEY, S., and G. W. IMBENS. "Chapter 3: The Econometrics of Randomized Experiments," in *Handbook of Economic Field Experiments*, Vol. 1, edited by A. V. Banerjee and E. Duflo. 2017: 73–140.
- AZMAT, G., and N. IRIBERRI. "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students." *Journal of Public Economics* 94 (2010): 435–52. https://doi.org/10.1016/j.jpubeco.2010.04.001
- BANKER, R. D.; H. CHANG; and M. J. PIZZINI. "The Balanced Scorecard: Judgmental Effects of Performance Measures Linked to Strategy." *The Accounting Review* 79 (2004): 1–23. https: //doi.org/10.2308/accr.2004.79.1.1
- BARANKAY, I. "Rank Incentives: Evidence from a Randomized Workplace Experiment." Wharton School, University of Pennsylvania, 2012. Available at http://repository.upenn.edu/ bepp_papers/75
- BLADER, S.; C. GARTENBERG; and A. PRAT. "The Contingent Effect of Management Practices." *Review of Economic Studies* 87 (2020): 721–49. https://doi.org/10.1093/restud/rdz034
- BLANES I VIDAL, J., and M. NOSSOL "Tournaments Without Prizes: Evidence from Personnel Records." Management Science 57 (2011): 1721–36. http://www.jstor.org/stable/41261969
- BRICKMAN, P., and R. BULMAN (1977). "Perspectives Pleasure and Pain in Social Comparison," in Social Comparison Process: Theoretical and Empirical Perspectives. Hemisphere.
- BULLARD, B. Style and Statistics: The Art of Retail Analytics. The Wiley and SAS Business Series, 2016.
- BURLIG, F.; P. LOUIS; and M. WOERMAN. "Panel Data and Experimental Design." Journal of Development Economics 144 (2020): 1–15. https://doi.org/10.3386/w26250
- BURSZTYN, L., and R. JENSEN. "How Does Peer Pressure Affect Educational Investments?" The Quarterly Journal of Economics 130 (2015): 1329–67. https://doi.org/10.1093/qje/qjv021
- CARDINAELS, E. "The Interplay Between Cost Accounting Knowledge and Presentation Formats in Cost-Based Decision-Making." Accounting, Organizations and Society 33: (2008) 582– 602. https://doi.org/10.1016/j.aos.2007.06.003
- CARDINAELS, E., and P. M. G. VAN VEEN-DIRKS. "Financial versus Non-Financial Information: The Impact of Information Organization and Presentation in a Balanced Score Card." Accounting, Organizations and Society 35 (2010): 565–78. https://doi.org/10.1016/j.aos.2010. 05.003
- CASAS-ARCE, P.; M. M. CHENG; I. GRABNER; and S. MODELL. "Managerial Accounting for Decision-Making and Planning." *Journal of Management Accounting Research* 34 (2022): 1–7. https://doi.org/10.2308/jmar-10784
- CHEN, Y.; J. JERMIAS; and T. PANGGABEAN. "The Role of Visual Attention in the Managerial Judgement of Balanced-Scorecard Performance Evaluation: Insights from Using an Eye-Tracking Device." *Journal of Accounting Research* 54 (2016): 113–46. https://doi.org/10. 1111/1475-679X.12102
- EYRING, H., and V. G. NARAYANAN "Performance Effects of Setting a High Reference Point for Peer-Performance Comparison." *Journal of Accounting Research* 56 (2018): 581–615. https: //doi.org/10.1111/1475-679X.12199
- FLOYD, E., and J. A. LIST "Using Field Experiments in Accounting and Finance." Journal of Accounting Research 54 (2016): 437–75. https://doi.org/10.1111/1475-679X.12113

- FRIEBEL, G.; M. HEINZ; M. KRUEGER; and N. ZUBANOV. "Team Incentives and Performance: Evidence from a Retail Chain." American Economic Review 107 (2017): 2168–203. https:// doi.org/10.1257/aer.20160788
- GILL, D.; Z. KISSOVÁ; J. LEE; and V. PROWSE. "First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision." *Management Science* 65 (2019): 494–507. https://doi.org/10.1287/mnsc.2017.2907
- HANNAN, R. L.; G. P. MCPHEE; A. H. NEWMAN; and I. D. TAFKOV. "The Effect of Relative Performance Information on Performance and Effort Allocation in a Multi-Task Environment." *The Accounting Review* 88 (2013): 553–75. https://doi.org/10.2308/accr-50312
- HANNAN, R. L.; G. P. MCPHEE; A. H. NEWMAN; and I. D. TAFKOV. "The Informativeness of Relative Performance Information and its Effect on Effort Allocation in a Multitask Environment." *Contemporary Accounting Research* 36 (2019): 1607–33. https://doi.org/10.1111/1911-3846.12482
- HOEGL, M., and H. G. GEMUENDEN. "Teamwork Quality and the Success of Innovative Projects: A Theoretical Concept and Empirical Evidence." *Organization Science* 12 (2001): 435–49. https://doi.org/10.1287/orsc.12.4.435.10635
- KNAUER, T.; F. SOMMER; and A. WÖHRMANN. "What Is a Good Rank? The Effort and Performance Effects of Adding Performance Category Labels to Relative Performance Information." *Contemporary Accounting Research* 38 (2020): 839–66. https://doi.org/10.1111/1911-3846.12657
- KNIPPENBERG, V. "Work Motivation and Performance: A Social Identity Perspective." Applied Psychology 49 (2000): 357–71.
- KOHLER, M.; M. D. MAHLENDORF; M. SEITER; and T. VOGELSANG. "The Polarizing Performance Effect of Private Social Comparison Information." Working paper, 2023.
- KRAMER, S.; V. S. MAAS; and M. v. RINSUM. "Relative Performance Information, Rank Ordering, and Employee Performance: A Research Note." *Management Accounting Research* 33: (2016) 16–24. https://doi.org/10.1016/j.mar.2016.03.004
- KUHNEN, C. M., and A. TYMULA. "Feedback, Self-Esteem, and Performance in Organizations." Management Science 58 (2012): 94–113. https://doi.org/10.1287/mnsc.1110.1379
- Lu, J. "Limited Attention: Implications for Financial Reporting." *Journal of Accounting Research* 60 (2022): 1991–2027. https://doi.org/10.1111/1475-679X.12432
- LUCA, D. L., and R. COLE. "Evaluation Technical Assistance Brief for OAH Teenage Pregnancy Prevention Grantees." HHS Office of Adolescent Health, 2017. Available at https://opa. hhs.gov/sites/default/files/2020-07/tppeval-tabrief9.pdf
- MAHLENDORF, M. D.; F. KLEINSCHMIT; and P. PEREGO. "Relational Effects of Relative Performance Information: The Role of Professional Identity." Accounting, Organizations and Society 39 (2014): 331–47. https://doi.org/10.1016/j.aos.2014.05.001. SSRN Working Paper.
- MAHLENDORF, M. D., and T. VOGELSANG. "Performance Measures and Incentives: Evidence from Field Experiments in Firms." (2023). https://doi.org/10.2139/ssrn.4407742
- MANTHEI, K.; D. SLIWKA; and T. VOGELSANG. "Performance Pay and Prior Learning: Evidence from a Retail Chain." *Management Science* 67 (2021): 6998–7022. https://doi.org/10.1287/ mnsc.2020.3825
- MANTHEI, K.; D. SLIWKA; and T. VOGELSANG. "Information, Incentives, and Attention: A Field Experiment on the Interaction of Management Controls." *The Accounting Review* 98 (2023): 455–79. https://doi.org/10.2308/TAR-2019-1065
- McGREGOR, J. "The Struggle to Measure Performance." Bloomberg Businessweek, 2006. Available at https://www.bloomberg.com/news/articles/2006-01-08/the-struggle-to-measureperformance
- PRENDERGAST, C. "The Provision of Incentives in Firms." Journal of Economic Literature 37 (1999): 7–63. https://doi.org/10.1257/jel.37.1.7
- ŞABAC, F., and J. YOO. "Performance Measure Aggregation in Multi-Task Agencies." Contemporary Accounting Research 35 (2018): 716–33. https://doi.org/10.1111/1911-3846.12418
- SIMONS, R., and A. DÁVILA. "How Top Managers Use the Entrepreneurial Gap to Drive Strategic Change." *European Accounting Review* 30 (2021): 583–609. https://doi.org/10.1080/ 09638180.2020.1792959

- SOFOKLIS, G., and R. MEGALOKONOMOU. "Knowing Who You Actually Are: The Effect of Feedback on Short- and Long-Term Outcomes." *Journal of Economic Behavior and Organization* 183: (2021) 589–615. https://doi.org/10.1016/j.jebo.2021.01.013
- SONG, H.; A. L. TUCKER; K. L. MURRELL; and D. R. VINSONC. "Closing the Productivity Gap: Improving Worker Productivity Through Public Relative Performance Feedback and Validation of Best Practices." *Management Science* 64 (2018): 2628–49. https://doi.org/10.1287/ mnsc.2017.2745
- STEELE, C. M. "The Psychology of Self-Affirmation: Sustaining the Integrity of the Self." Advances in Experimental Social Psychology 21 (1988): 261–302. https://doi.org/10.1016/S0065-2601(08)60229-4
- SULS, J., and L. WHEELER (Eds.). *Handbook of Social Comparison: Theory and Research*. Kluwer Academic Publishers, 2000. https://doi.org/10.1007/978-1-4615-4237-7.
- TAFKOV, I. D. "Private and Public Relative Performance Information under Different Compensation Contracts." *The Accounting Review* 88 (2013): 327–50. https://doi.org/10.2308/accr-50292
- WILLIAM, M. C.; H. JIAPENG; L. WENWEI, S. CHENGDAO; W. DI; and Z. YANG. "IS There a Confidence Interval for That? A Critical Examination of Null Outcome Reporting in Accounting Research." *Behavioral Research in Accounting* 34 (2022): 43–72. https://doi.org/10.2308/ BRIA-2020-033