Is This Really Kneaded? Ask the Manager! A Large-scale Trial on the Effects of Paperwork Reduction*

Guido Friebel[†] Matthias Heinz[‡] Mitchell Hoffman[§]

Tobias Kretschmer[¶] Nick Zubanov[‖]

September 2022

Abstract

In a large German bakery chain, many workers report negative perceptions of paperwork. Randomly removing two of the most onerous paperwork duties in half of stores, sales increase by almost 3%. Effects appear not to come from workers having more time for production. Rather, most effects are indirect, with workers intrinsically valuing that the firm reduced their paperwork. Beneficial effects are fully concentrated in stores where regional managers predict that the treatment will be effective, reflecting substantial heterogeneity in returns that is well-understood by managers. Sales impacts are also smaller in larger stores, reflecting that some paperwork helps coordinate production. As a result of the RCT, the firm implemented firmwide reductions in paperwork, eliminating paperwork that workers regard as demeaning, but keeping paperwork that helps coordinate production.

JEL Classifications: M51, M52, D23
Keywords: Management practices; organizational economics; time use; respect

*We thank the study firm and its management for their enthusiastic participation in this collaboration. We thank Charlie Brown, Axel Ockenfels, Andrea Prat, and Melanie Wasserman, as well as seminar participants at Rotman and WashU Olin for their comments and suggestions. Financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC 2126/1– 390838866) and SSHRC is gratefully acknowledged. The experiment was pre-registered on 04/14/2021 with the AEA RCT registry under ID AEARCTR-0007550. IRB approval was received from the University of Cologne. The Worker Council approved the project and was involved in all steps.

[†]Goethe University of Frankfurt and CEPR and IZA
[‡]University of Cologne and CEPR
[§]U. Toronto Rotman School of Management and NBER
[¶]LMU Munich
[‖]University of Konstanz and IZA

Firms differ markedly in their management practices and such differences are widely held to help explain performance differences across firms (Ichniowski *et al.*, 1997; Ichniowski & Shaw, 1999; Syverson, 2011). One key aspect of management is monitoring, broadly defined as keeping a close watch on the production process. Growing research, including randomized controlled trials (RCTs), emphasizes that improving monitoring can enhance performance (Duflo *et al.*, 2012; Jackson & Schneider, 2015; Bandiera *et al.*, 2021), and aspects of monitoring are scored in the World Management Survey (Bloom & Van Reenen, 2007). However, increased monitoring need not always be beneficial. Separate from the direct cost of monitoring technology, monitoring often takes time, both for monitors and the people being monitored. Furthermore, employees may dislike being monitored, not only because it prevents them from slacking off, but also because it may be a signal of mistrust (Falk & Kosfeld, 2006). Indeed, excessive use of monitoring could even lead to a negative company culture based on fear and mistrust.

Consider paperwork, a common form of monitoring. Paperwork is pervasive at work, though often disliked (Strausz, 2006), and the degree of paperwork varies substantially across organizations.[1] Take the example of a small-scale travel expense, say $50, for an academic. Some universities issue credit cards so that people can pay for small expenses with no receipts. Many universities require receipts, and at least one requires both receipts and credit card statements to corroborate payment. Of course, firms' intended goal of paperwork is not to create pointless work, but to achieve a beneficial end, such as preventing employee malfeasance, to help coordinate people's work, or to help workers remember to do things (Gawande, 2010).

As far as we know, we provide the first RCT on paperwork reduction in any field. Our research partner is a major German bakery chain with 150 stores, over 2,000 workers, and over €100m of annual revenue. The firm is family-run and prior to our intervention was using extensive paperwork in many aspects of production. Workers needed to record extensive information, not only about their products (e.g., when they took bread out of the oven), but also on their interactions with customers, such as whether they smiled. Drawing on a deep collaboration with the firm and top management, we conducted extensive pre-RCT interviews and surveys within the firm, and discovered that several aspects of paperwork were particularly resented by workers. Management agreed to randomly remove two of the most onerous paperwork duties in half of stores.

The RCT is grounded in a simple conceptual framework of monitoring, as laid out in

[1] Dislike of paperwork is common in popular culture. A famous example is the 1999 movie *Office Space*, where the main character, a software engineer, spends substantial time on TPS reports. For recent general discussion on aversion to monitoring, see https://www.nytimes.com/interactive/2022/08/14/business/worker-productivity-tracking.html?smid=nytcore-ios-share&referringSource=articleShare.

Section 1. Paperwork helps firms address moral hazard problems, coordinate production, and remind workers of tasks. However, paperwork also entails costs, both in terms of time and in terms of signaling mistrust.

As detailed in Section 2, the bakery chain we study represents an ideal setting for our RCT. First, the sample is large, with 148 stores and thousands of workers. Second, we have access to highly granular administrative data, coupled with the ability to conduct high-quality, detailed surveys. The administrative data cover detailed aspects of sales, customers, and orders hour by hour, which is critical for examining how workers and managers are using their time and how they substitute time on paperwork to other tasks. Because of our deep collaboration, the surveys we conduct have very high response rates, as well as in-depth open-ended questions, which are critical for understanding mechanisms. Unusually, we survey not only store employees and managers, but also regional managers (the bosses of store managers) and sales directors (the bosses of regional managers) in detail, and have them make predictions about in what stores the RCT will be most successful. Third, bakery jobs like the ones we broadly representative of many less-skilled jobs around the world.

Focusing first on the overall effects of the RCT, we estimate that removing paperwork increases sales by 3%. The impact on sales is similar during busy and less busy times. While one may be concerned that removing monitoring would lead to wasted food, increases in employee misbehavior, or coordination failures, we observe no negative impact on shrinkage (a joint measure of food waste and worker stealing). We also observe no negative impact on mystery shopping scores. Our bakery firm has relatively low attrition, and there is no overall impact of the treatment on attrition. However, there are disparate effects by type of worker, with positive effects on attrition on mini-jobbers, the unskilled part-time workers who may benefit from structure and checklists, as well as negative effects on qualified workers and especially on managers.

Our initial discussions with regional managers strongly suggested, however, that the impacts of the RCT on outcomes would be highly heterogeneous across stores. In our pre-RCT survey of regional managers, managers predicted that in about half of their stores the treatment would be effective, and in the other half they would not. Thus, in our RCT pre-registration, we focused strongly on this aspect of heterogeneity. Splitting the sample based on whether regional managers predicted the store would work, we observe vast differences in the results. Among stores where the RCT was predicted to be successful, we find that removing paperwork increases sales by 5%. There are broad-based improvements in store operations, with round-the-clock improvements in sales, statistically increases in customers, and a decrease in shrinkage. In contrast, in stores where the treatment was not predicted to work, the impact on both store-level outcomes and employee attrition is zero. If anything,

mystery shopping scores are slightly down, though the impact is not statistically significant.

To better understand these effects, we dig into the free text of regional managers' responses on why the treatment would work in particular stores. Among stores where regional managers predicted the treatment will work, in about one-third of cases, regional managers explicitly mention something about workers enjoying the reduction in paperwork, consistent with a utility cost to excessive paperwork. In about two-thirds of cases, regional managers mention something about the absence of problems, consistent with traditional views of monitoring to help detect and avoid problems.

The firm was very pleased with the results of the RCT. Unlike past interventions in the literature, our treatment was taking something away (instead of adding something) so the direct cost to implement the RCT was very low. For minimal cost, the firm received a sustained increase in sales, as well as a reduction in manager turnover. Therefore, the firm decided to implement paperwork reduction firmwide. However, while the RCT involved eliminating two paperwork duties, the firm decided to restore the daily protocol in the firmwide rollout even though the operational checklist was eliminated.

Our paper contributes to several literatures. First, it contributes to work in personnel and organizational economics, as well as social science more general, on the returns to checklists and monitoring.[2] Most influentially, the physician Atul Gawande (2010) summarizes studies and in-person observations from a number of domains, including those of surgeons (see Ko *et al.* (2011) for a review), airline pilots, and investors, to argue that checklists can have profound positive organizational consequences. Our findings show that the returns to monitoring need not be positive, as we estimate sizable positive benefits of removing checklists. The central reason, we believe, is the presence of indirect costs of monitoring. Using lab experiments with assigned roles, Falk & Kosfeld (2006) show that workers react negatively and often choose low effort when being controlled by the manager. Our paper suggests that such insights extend into the field as well, and we offer a framework that rationalizes why checklists may be good for some tasks, but bad for others.

In economics field experiments on monitoring, most related to ours is a seminal paper by Nagin *et al.* (2002). Nagin *et al.* (2002) consider a field experiment where a call-center company exogenously varies its monitoring rate in some call-centers. They show that increasing the monitoring rate leads to a decrease in suspected bad calls, but that a certain share of workers do not appear to respond to the additional monitoring, due to a belief that workers should behave in an appropriate manner. However, the fundamental psychology at

[2]In economics, experiments (lab and field) showing impacts of monitoring include Duflo *et al.* (2012); Jackson & Schneider (2015); Dickinson & Villeval (2008); Falk & Kosfeld (2006); Bandiera *et al.* (2021); Kelley *et al.* (2021). In most of these studies, monitoring is added instead of taken away. There are also many observational studies, especially in the trucking industry (Hubbard, 2000, 2003).

the heart of our study is different. Our study is about workers drawing negative signals from the firm due to the presence of excessive monitoring instead of workers choosing to behave in an appropriate manner despite the presence of monitoring. Despite this difference, we believe the results are highly complementary, and suggest important augmentations of the impact of monitoring technologies beyond their traditional role in detecting malfeasance (Holmstrom, 1979).[3]

Second, our paper contributes to work in personnel economics on the heterogeneous returns to management practices and on the impact of managers. In the midst of substantial work on the general importance of management practices (Bloom *et al.*, 2012, 2019), growing research emphasizes that management practices are complementary to one another (Milgrom & Roberts, 1990; Ichniowski *et al.*, 1997), and that their impact may be contingent on other factors within an organization (Blader *et al.*, 2020). We show that there is substantial heterogeneity in the return to a management practice, namely, paperwork reduction, based on manager beliefs. Manager beliefs are somewhat correlated with some observable traits of stores, e.g., managers correctly predict that the treatment will be larger in smaller stores, but there is substantial predictiveness of manager beliefs beyond observable characteristics. A rich and growing literature examines what do non-CEO managers do and their impact (Lazear *et al.*, 2015; Friebel *et al.*, 2022; Frederiksen *et al.*, 2020; Hoffman & Tadelis, 2021), often emphasizing the role of managers in motivating and teaching employees. Our results suggest that an important value of managers is their private information (Aghion & Tirole, 1997).

Third, our paper makes a methodological contribution to RCTs. Beginning with DellaVigna & Pope (2018), growing work uses expert predictions for the purpose of examining how the results of an RCT compared to priors of experts, that is, to see to what extent a result is surprising or not (DellaVigna *et al.*, 2019). Rather than having experts predict the average results of the RCT (e.g., that the treatment will increase or decrease sales by a certain amount), our RCT has experts predict store by store whether the treatment will be effective in that particular store. We are not aware of other RCTs that use expert predictions in this manner, but we believe that this is a methodology that may be useful in other contexts.[4] We show that experts in our context have substantial knowledge about which units will be most affected. Our RCT also suggests a new role for subjective expectations.

[3] de Rochambeau (2020) shows that randomly monitoring Liberian truckers increases their effort, though there are some workers who reduce their output after being monitored. Hiring students to identify coins, Belot & Schröder (2016) show that randomly added monitoring can backfire on some dimensions of performance.

[4] For example, one could imagine asking doctors or patients to predict which individual patients will be responsive to a new drug, or to asking people in a village which villagers will benefit from increased credit.

1 Conceptual Framework

What is the impact of monitoring such as checklists on performance, and how would the impact of monitoring vary across stores? To address these questions, consider the decision of a firm whether or not to use monitoring.

The firm faces *problems*, which we define broadly. First, these can be memory problems, where people on a surgery team forget to take certain steps (Gawande, 2010) or it can be a bakery worker forgetting to put the doughnut at the correct angle. Second, and very importantly for us, these can be coordination problems, e.g., a bakery worker forgets to pass along to the next shift at what time the bread was made. Finally, these can also be problems of moral hazard where workers behave opportunistically (Nagin *et al.*, 2002). To keep things absolutely simple, we assume that problems occur exogenously with probability p, but the logic of our model can be easily extended to having workers choosing whether to behave opportunistically. When a problem occurs, the cost to the firm is k. Thus, without monitoring, firm profits are $-pk$.

Monitoring such as checklists helps the firm identify problems. The quality of monitoring is given by m, and represents the probability that a problem is detected and solved in full. Equally, one can assume that monitoring detects problems with 100% probability, but that only a share m of costs are recuperated. Using monitoring also involves direct cost, c, which can include the technology itself, but in our setting is primarily the time cost of filling out paperwork.

In addition, monitoring entails an indirect cost θ to firm performance. Many people seem to dislike being monitored, perhaps because it is intrinsically unpleasant to fill out paperwork but also because monitoring can be viewed as a sign of disrespect (Ellingsen & Johannesson, 2007, 2008). Ellingsen & Johannesson (2008) argue that workplace respect can be thought of in terms of second-order beliefs, i.e., a worker's belief about the firm's belief about whether she is altruistic or competent. Being respected can be important for firm performance, both because it makes workers more likely to stay with the firm (Friebel *et al.*, 2023) but also because it motivates them to work harder (Cai & Wang, 2022). Alternatively, being monitored could crowd out intrinsic motivation to work hard (Benabou & Tirole, 2003).

Therefore, the profits from monitoring are $-(1-m)pk - c - \theta$, and the firm chooses to monitor when $mpk - c - \theta > 0$. This expression allows us to characterize what are the stores where monitoring will have the largest positive effects. Specifically, monitoring is most useful in a firm that faces more frequent problems, when the memory technology can more reliably identify problems, and where the cost of those problems are higher.

This framework also raises the possibility that there could be substantial heterogeneity across stores within a firm in the returns to monitoring. Managers may know that some stores experience frequent coordination problems and thus likely benefit from monitoring. Stores may also vary in the disutility that workers receive from being monitored. Given that there are multiple factors affecting whether monitoring has positive effects, as well as that some factors (like frequency of coordination problems) is very difficult to measure in data, it is natural to ask regional managers to make predictions about whether a treatment will look.

2 Study Background

The study firm. The firm is one of the largest bakery chains in one densely populated region of Germany.[5] The firm is family-owned, as is typical for most bakery chains. The CEO is also the founder of the modern version of the bakery chain. The company has roughly 2,000 employees. Many of the top executives helped set up the chain with the CEO over the last 40 years. The firm has 150 stores, as well as one plant which produces raw products for the stores (e.g., unbaked bread which is baked in store ovens).

Why the firm did the RCT. Our collaboration with the firm arose at the beginning of 2020 when the firm became aware of a highly successful intervention in a Germany bakery chain involving three of this paper's coauthors (Friebel *et al.*, 2017). In initial exploration about potential projects, the authors came across a 2019 employee survey which showed widespread general dissatisfaction with paperwork at the firm. Based on this, we thought it would be promising to examine this issue in greater detail and rigor. The authors and firm formed a project team consisting of two of paper's authors, the head of the HR department, the head of the accounting/controlling department, multiple employees from those two departments, one sales director, and the head of the worker's council.

The project team identified all the 22 existing documentation duties in stores. In-depth in-person surveys by RAs were then conducted with 21 store managers and 18 workers in 22 randomly selected shops about beliefs regarding time use duties. Because of the control-oriented culture, the researchers were concerned that there would be issues of trust, so we pushed for the surveys to be done in-person, as we believed that we would get more truthful and accurate information this way compared to online surveys. To further cement trust, the RAs were driven to the stores by the head of the worker's council. The head introduced the RAs to the survey respondents, emphasizing that they could trust the RAs. For each of the

[5] In Germany, most bakery chains operate in particular regions.

22 duties, respondents were asked the following questions:[6]

1. to what extent the duty helps the company achieve its goals (1-10 scale)?
2. to what extent the duty helps avoid making mistakes (1-10 scale)?
3. how often do you fill out the paperwork duty each week?
4. how many minutes do you spend each time filling out the paperwork?

Results on the survey are provided in Figure 1. As can be seen, 5 duties stood out for having an unfortunate combination of relatively low value and higher time cost. 3 of these were seen as "holy cows" and not possible to remove either for political reasons or because they were seen as related to the unique selling proposition of the firm.[7] The remaining two duties were the **operational checklist** (*operative liste*) and the **daily protocol** (*tagesprotocol*). In a meeting in October 2020, the researchers presented analysis on these surveys and recommended removing these two duties.

We also observe that workers and managers in the in-depth interviews have similar average beliefs about how much time the paperwork duties take; these beliefs are correlated (correlation coefficient: 0.77) with the beliefs of the top management members from our project team, which we elicited in a project team meeting before conducting the in-dept interviews. This suggests that the data in the in-dept interviews are high quality and that people took the questions seriously. It also indicates that the top management was well aware of the time for paperwork duties.

[6] In the interviews in randomly selected stores, we only asked the questions for 21 documentation duties. The "missing" documentation duty is called "Einverständniserklärung Sonntagsarbeit" (=declaration of consent for working on Sundays): When employees have holidays on a Friday, it is not clear whether the weekend counts to the holidays or is already the start of the new working week. According to the top management of the firm, the weekend does not account towards the holidays (as long as workers do not have holidays on the upcoming Monday); according to the workers' council, the weekend is part of the holidays from the previous week. There was a big dispute between the firm and the workers' council about the case in the past. The compromise: Saturday counts as holidays and workers are not allowed to work on that day; Sunday counts as a holiday, workers are, however, allowed to work on that day and the hours are counted as overtime – however, workers have to explicitly declare that they are willing to work on the Sunday (i.e., sign the "Einverständniserklärung Sonntagsarbeit"). The top management and the workers' council also made a "work agreement" about the compromise – which is an Germany a legal binding agreement between the firm and its workers (i.e., the document is legally treated on the company level in the same way as a the German labor law). When we prepared the interviews, both the top management and the workers' council informed us that it is legally and politically impossible to drop this document. Also note that the document is rarely used (only if employees work on Sundays after their holidays) and it takes only one minute to sign it. Because of that, we did not include this documentation duty in our interviews.

[7] For example, one duty seen as having lower value and higher time is the so-called *goldbroetchen* or "golden roll," where every time a bakery bakes a load of rolls they need to send 5 rolls to the headquarters for potential examination or testing by top management. However, baked rolls and roll quality are considered essential to the company's unique selling proposition, so the golden roll duty could not be removed.

Given the extensive preparation by the research team into locating the least useful and most onerous paperwork duties, the company was very interested in running an RCT, and consider making a change. The company is no stranger to experimentation, and frequently engages in "pilots" with some of the shops (e.g., new products, marketing campaigns, shop design). Thus, the fact that there were significant changes in some shops would not have been as anomalous or bizarre, with no one thinking that the company never makes changes.

Broadly speaking, within top management, there were two broad "schools of thought" regarding the company's paperwork duties. One group emphasized the benefits of monitoring, pointing out the importance of *struktur* (or structure) for workers, especially given the firm has 150 stores, and stores can't be monitored personally by top management all the time. The other group emphasized the costs of paperwork duties, both the time involved and the idea that monitoring is a signal of disrespect. Thus, the firm's pre-RCT debates on paperwork paralleled the tradeoffs emphasized in the academic literature. The executives in the pro-structure school of thought were the ones who initially introduced all the documentation duties to the firm. Thus, these executives have much longer tenure than the executives emphasizing the costs of paperwork duties.

Operational checklist. The operational checklist is a detailed check list such as, I smiled at the customers, I put the rolls at the right place in the shelves, I put the cinnamon sweets in the right way on the counter, I put the sugar on the Berliner doughnut in the right shape, I finished the old package of coffee beans before starting a new package of coffee beans, I know about the covid restrictions. As seen in Figure 2, which provides the operational checklist from right before the RCT, it is a constant reminder for workers about how they are supposed to do their jobs. In our initial focus groups, many workers view the list as somewhat insulting. Employees are required to sign the operational checklist every day. Workers complete the operational checklist at different points in the day.

A key point about the operational checklist is that stores are provided the same information in the operational checklist in the form of a weekly newsletter. For example, the newsletter already tells the stores about the correct placement of Berliner doughnuts. Thus, the operational checklist is reminding workers and requesting signatures on behavior that they have already been reminded about. In short, workers are constantly being reminded how to do their daily job, including in the newsletter, and then the daily protocol reminds them of what they have already been reminded of.

Daily protocol. The second duty we study is the daily protocol, where you write down all the things that happened during the day. This includes how much money is in the cash register, how much sales taken in, and whether workers would like to pass along this information to the daily shift (this last point seemed especially appealing in bigger stores). In

contrast to the operational checklist, some workers find value in the daily protocol. Managers initially thought that without the operational checklist, stores would experience significant operational problems, and that workers would not follow the guidelines of the company (e.g., employees would forget to keep the shelves clean and smiling at the customer). In our survey data, workers report varying amounts of time to do the daily protocol. Some workers report doing it in 5 minutes, whereas others take as much as 60 minutes. Workers do the daily protocol at the end of their shifts.

An important point about both the operational checklist and daily protocol list is that the documents are rarely examined by corporate headquarters, and workers perceive that they are never looked at by headquarters.

RCT setup with regional managers. The RCT treatment consists of removing two forms of paperwork in treatment stores. Regional managers and the sales managers were invited to a meeting on February 16, 2021 with top executives and the research team. Regional managers were informed that there would be a 6-month RCT and were provided detailed guidelines about the RCT. Regional managers were also informed that surveys would be administered, and were given the opportunity to ask questions.

In the meeting, several regional managers immediately expressed strong opinions on in which of their stores the treatment would be effective. This strongly suggested the importance to us of understanding heterogeneous treatment effects, and it seemed that regional managers may possess some strong local knowledge on this heterogeneity. Thus, in March 2021, before knowing which stores were in control or treatment, regional managers also made predictions by phone about in which stores the treatment would be effective (exact wording in Appendix C.1). We interviewed all 15 regional managers (100% response rate), with all interviews conducted by one coauthor.[8] We motivated the phone call to the regional managers using the fact that there was significant heterogeneity in managers' predictions (and the rationales) for whether the treatment would work. To make the predictions as natural as possible, we asked regional managers for verbal responses, which we then translate on our own into a numerical response of whether it will work. For almost all of the responses, there is little ambiguity about opinions, as we detail in the Appendix. No incentives are used for this prediction because it is a subjective one.[9]

[8]It was important in our view to have the interviews conducted by a coauthor since it would not be respectful or appropriate to have such senior managers conducted by a research assistant. We also thought we would get more serious and complete responses if a chaired German professor called instead of a research assistant.

[9]Furthermore, we do not use incentives to avoid any reward for regional managers so as to influence or manipulate outcomes in stores, and it also seemed most natural to us avoid an incentive for these higher-profile managers. Avoiding incentives also reduced the salience the prediction, e.g., to make it something that would be out of the ordinary and thus something to linger in managers' minds in the future. Finally, in a

The three sales directors did not make formal predictions, but we learned in the course of our discussions about which of the sales directors were most optimistic about the treatment. One sales director was optimistic, one was pessimistic, and one was in between, and as pre-registered, we examine heterogeneity

RCT setup with store managers and workers. Store managers and workers were informed about the RCT via the firm's weekly newsletter. The information came to a message on the store intranet on Tuesday April 6, 2021 (after the Easter holiday) and also on paper form in the bundle of papers for the weekly documentation duties. In contrast to regional managers, workers and managers were not informed that there was an RCT or that the change would last for a certain period of time. Workers and managers in the treatment group indicated full awareness of the treatment. This is natural given that the RCT removed paperwork duties that were an important part of the normal job.

The message in the firm's weekly newsletter informing treatment stores about the change came from the firm's COO, who is the son of the CEO/owner. Having the message come from the COO gives credibility and seriousness to the change. As seen in Figure X, the message emphasizes two things, paralleling our hypotheses on direct and indirect effects. First, it emphasizes how the company is trusting workers. Second, it emphasizes the extra time, and that workers should use the extra time for customers and colleagues. One reaction to this is that it might seem that workers are being "led" to think a certain way. However, it would be extremely artificial and unattractive to a company to make a large change like removing significant documentation duties without explaining why. Moreover, even if workers being led, that couldn't explain.

We took great pains to ensure that the RCT was carried out as planned. Store paperwork is delivered every week to stores in a bundle. We sent an RA to monitor that the paperwork bundles delivered to treatment stores did not contain the operational checklist or daily protocol, but that control stores did.[10] We also called regional managers, the head of HR, and one sales director in May 2021 to ensure that the treatments were being carried out as planned, and there were no issues with implementation.

RCT timing. The experiment began on April 6, 2021. Paperwork was removed in treatment stores. The authors presented the results to the firm in December 2021. Given the success of the RCT, the firm immediately decided to roll out the treatment to control stores, and this began at the end of January 2022. In the rollout, the operational checklist

recent literature review, Haaland *et al.* (2022) that incentives are not needed to accurately elicit beliefs, and can sometimes make elicitation worse. Hoffman & Burks (2020) show that randomly providing workers with incentives to accurately predict their productivity has no effect on beliefs, suggesting that belief incentives are not important for eliciting beliefs.

[10] After 6 weeks, the firm asked if the RA could come only every couple weeks, and we granted this.

was kept removed from the company. However, the daily protocol was introduced, given that some workers found it useful and less onerous. The RCT received IRB approval from the University of Cologne, and it was also approved by the Worker Council of the study firm.

The RCT was registered on the AEA RCT Registry on April 14, 2021. Our analyses closely follow the registration. Based on theory and our interactions with the companies, we pre-registered that there would be treatment effect heterogeneity according to team size, team tenure, and regional manager predictions.

Data. We use administrative data from the firm to create two main panel datasets. First, we create a store-level panel, exploiting detailed hourly data on sales by store. The store-level dataset also includes information on mystery shoppers. Second, we create a worker-month panel covering worker absence, worker attrition.

The pre-treatment store manager survey was conducted in March 2021. It was a phone survey conducted by RAs and the response rate was roughly 95%, with N=135. The pre-treatment regional manager survey was conducted after regional managers the existence of the RCT, but before the knew which stores were in the treatment group. The main purpose of this survey was to assess regional manager beliefs about in which stores the treatment would work. The during-RCT store manager survey was conducted in November 2021, also conducted by phone.

Finally, there was a during-RCT worker survey, conducted with pen and paper in the stores, in October 2021. This survey was conducted using a large number of RAs who personally visited the stores and collected the questionnaires.

Randomization. We conducted a stratified randomization using 4 dimensions of stratification: pre-RCT head count (above or below median), pre-RCT sales (above or below median), pre-RCT store ranking in the firm's performance league (above or below median, with this variable described more in the Data Appendix), and town (8 towns). This gives us 64 strata. Randomization was conducted using "randtreat" in Stata. As seen in Table 1 below, we observe strong balance across various characteristics, both stratified and unstratified.

3 Main Results: The Impact of the Treatment on Store-level Outcomes and Employee Attrition

To estimate the impact of the treatments, we consider ANCOVA specifications following Bruhn & McKenzie (2009). We use OLS models where we control for the mean of the dependent variable in the pre-RCT period, as well as store-month fixed effects for the RCT

period. We have also estimated models controlling for the variables used in stratification and our conclusions are the same. Throughout the paper, standard errors are clustered by store, reflecting the level of randomization.

Store-level outcomes. Panel A of Table 2 shows that the treatment boosted sales. Overall sales went up by 2.8%, statistically significant at the 10% level. Sales increased both at the busiest part of the day for bakeries (7am to 2pm) and in the less-busy sections. One concern with removing paperwork is that it could lead to a decrease in product quality, a decrease in employee effort, and an increase in employee misbehavior. However, we see little evidence for that. Shrinkage and the mystery score are both unchanged, and we can reject that there are significant negative effects.

Panels B and C separate the treatment effect by regional manager expectations, showing that the treatment effect is much stronger in stores where regional managers expected the treatment to be beneficial. In stores where managers expected the treatment to be beneficial, sales increase by 5%, statistically significant at the 5% level, with similar increases among busy and slow sales. The number of customers increases by 4.9%, and shrinkage—a combination of wasted product and theft—goes down 3.4%. In contrast, for stores where the treatment is not predicted to work, the effects on sales are zero and shrinkage *increases* by 3.6%, though this latter difference is not statistically significant.

The final row in Table 2 shows p-values testing the differences between stores where regional managers predict the treatment to work. We see that the difference in shrinkage effects is statistically significant, as are the differences in effects on sales and customers. We show two-sided p-values so as to be maximally conservative, though one can very easily argue that one-sided p-values are more appropriate given the explicitly one-sided prediction of store managers (i.e., dividing stores in the ones where the treatment will work and ones where it will not work).

In addition to show ANCOVA effects for the entire RCT period, it is useful to show effects over time. Figure 4 shows the impact of the treatment over time in stores where the treatment is predicted to work using an event study framework. We use store fixed effects and focus on the interaction of treatment status with time from the start of the RCT. The impact on sales occurs in the first quarter of treatment, consistent with the large distaste that many workers and managers at the firm expressed toward paperwork. The effect seems slightly lower in the second quarter of the RCT before recovering in the third and final quarter. However, one is not able to reject that the treatment effect is constant throughout the RCT period. Even in the last quarter of the RCT, coming 6-9 months after the treatment was introduced, paperwork reduction increases sales by 3.4%. The effect is especially persistent for sales during busy periods.

Attrition. Table 3 examines effects of the treatment on attrition. Results for attrition are relatively imprecise as there is not too much attrition at the firm. There is no overall effect on attrition, but this masks substantial heterogeneity by worker type. Among minijobbers, attrition increases by a statistically insignificant 0.8 percentage points per month, which is an increase of roughly 20%. However, among qualified workers (i.e., workers who already did an apprentice training), attrition decreases by 0.4 percentage points per month. The firm is much more focused on the attrition of trained workers relative to untrained workers. Among managers, attrition decreases by 1.15 percentage points per month, a reduction of roughly 60%. This decrease is statistically significant at the 10% level.

Panels B and C show that the reduction in managerial attrition is entirely driven by stores where the treatment is predicted to be successful. In stores where the treatment is expected to work, we observe a reduction in attrition of almost 90%. In contrast, in stores where the treatment is not expected to work, the effect is zero. A p-value on this interaction term is significant at the 5% level.

Why are there large effects on manager attrition but not worker attrition? One likely reason is that the costs of paperwork are especially strong for managers, particularly in the case of the daily protocol. Managers spend almost an hour per week completing the daily protocol, whereas for workers the required duration is closer to half an hour. In pre-RCT focus groups and discussion with the firm, there was a feeling that paperwork was preventing managers from doing some high-value activities, such as mentoring and teaching workers. It is also possible that utility costs of monitoring are especially bothersome for managers. Managers are supposed to act as leaders and monitors in the store. When the firm does extensive paperwork on top of this, the firm is communicating that it does not trust the manager to perform these functions by himself or herself.

Magnitudes. How should we think about the magnitudes of the estimates? In a study in another bakery chain, Friebel *et al.* (2017) find that providing a team performance bonus led to an increase in sales and customers by 3%. Thus, the overall effect of the impact of reducing paperwork is similar to the impact of providing a team performance. However, in stores where regional managers predict that the treatment will be successful, our estimated magnitude is 70% larger. Our treatment is much more cost-effective and profitable, as the treatment in Friebel *et al.* (2017) involves an increase of wages of 2.2%, whereas compensation is kept constant in our RCT. The seminal monitoring RCT by Nagin *et al.* (2002) look at effects on suspicious calls, but do not have data on sales.

Another study of a particular management practice is a seminal RCT on work from home by Bloom *et al.* (2014). This study finds that working from home led to a 4% increase in calls per minute, which is also similar to our effect on sales. Bloom *et al.* (2014) also find

that work from home reduces attrition by half, which is broadly similar to the attrition effect we observe on managers. However, the effect we observe in stores in which the treatment is predicted to work is larger (though with a large standard error, meaning we cannot reject that the effect would be at half, though we can reject that the effect is zero).

Having a large attrition reduction concentrated among managers is also seen in a recent RCT by Alan *et al.* (2023). Working with Turkish firms, the authors examine the impact of a module by a consulting company designed to improve the relational atmosphere in the workplace. They find that this module reduces manager attrition by roughly 80% while having much smaller impacts on worker attrition. The impact of paperwork reduction on manager attrition is thus broadly similar to the effect of a workplace relational module.

In sum, the effect of removing two onerous paperwork duties in our setting leads to treatment effects on the order of some of the most promising and highly regarded past management interventions. At the same time, we believe that our effect sizes are very plausible. Some readers may be surprised that removing paperwork has such quantitatively substantial effects. It is critical to remember that the paperwork removed was identified by workers and managers as being more onerous, with relatively low value and high time cost.

3.1 Mechanisms for the Regional Manager Predictions

Why are regional manager expectations predictive of the treatment effect? What is the rationale for their predictions, both positive and negative? To address this question, we use the raw text from regional managers pre-RCT predictions. The text of regional manager predictions is provided in Appendix Tables A1 and A2.

Looking through the responses, there are two salient features of text responses for stores where regional managers predicted that the treatment would work. First, in many cases, regional managers mention that workers will enjoy having less paperwork. For one store the regional manager said that workers "Would be very happy about less bureaucracy, less work as a result, do not like to work with notes and strict rules." This explanation would fall under the utility cost of monitoring described in Section 1. Second, in many cases, regional managers talked about how teams would be unlikely to face problems, especially because the team already had good communication. An example prediction is that one store "Could live without bureaucracy, very communicative branch management." Some prediction mention both that reducing paperwork will be good for worker utility and that there are no anticipated problems. For example, one manager predicted that the "Team will be glad when operational list is gone. No problems expected. Will work out!"

Table 5 summarizes key facts about regional manager predictions. Among the stores

where regional managers believe the treatment will be successful, in 32% of predictions, regional managers mention something about paperwork reduction benefiting worker utility. Likewise, in 66% of predictions, regional managers mention something related to ability to overcome problems. Thus, regional manager predictions strongly support both the traditional economic view of monitoring as a way of addressing problems (Holmstrom, 1979; Halac & Prat, 2016), as well as theories emphasizing the utility costs of monitoring (Falk & Kosfeld, 2006).

A natural concern in interpreting the results on regional manager predictions is whether results could be due to managers behaving differently in treatment vs. control stores. However, in the predictions, no regional manager said anything about an intent to behave differently in treatment vs. control stores, such as by visiting treatment stores more often. A different concern is that regional managers might have private information not about the efficacy of treatment, but rather about the coming of external shocks to stores (e.g., there will be a large festival next to a store in the coming months). However, no regional managers said anything in their prediction about external shocks.

3.2 Direct and Indirect Effects of Paperwork Reduction

Separate from regional manager predictions, what drives the improvements in store performance that we observe, as well as the reductions in manager attrition? Are people using the extra time that they have to perform other tasks, which we can think of as the direct effect of paperwork reduction? Or is there some other mechanism such as increased happiness, trust, or respect? The regional manager predictions indicate that at least for some stores, regional managers believe that indirect effects will be present, believing that removing paperwork will make workers happier.

Table 6 examines heterogeneity in the overall treatment effect on sales based on the amount of time that stores spend on the daily protocol in the pre-RCT period. As seen by the key interaction term, there is no evidence that the treatment effect on sales varies with time spent on the daily protocol. Rather than looking at the quantity of time, one can instead focus on when stores tend to do the daily protocol in the pre-RCT period. We find no evidence that the treatment effect is larger during the time periods when stores generally do the daily protocol. Recall that the daily protocol takes more time compared to the operational checklist.

These two pieces of evidence fail to support direct effects of the treatment, i.e., that paperwork reduction increases sales by allocating extra time to other activities. One additional piece of evidence in favor of indirect effects comes via the firmwide rollout, which we

discuss shortly below.

3.3 Threats to Validity

Hawthorne effects. A concern in any RCT is whether subjects could alter their behavior in order to please the researchers (Levitt & List, 2011). However, in our setting, workers and store managers did not know that they were part of an RCT, so there is no awareness of needing to please a researcher.

Contemporaneous policy changes. Another concern in any RCT is the presence of contemporaneous policy changes. However, this was not the case in our firm.

Multiple hypothesis testing. In a study addressing multiple outcomes and heterogeneous treatment effects, one worries that treatment effects could be spurious due to multiple hypothesis testing. The main way that we address this point is through the rigorous pre-registration of our RCT. Our main outcomes are listed in the pre-registration before the RCT began, and we explicitly say that our primary outcome is store sales. In addition, we explicitly say that our heterogeneity analysis will focus on heterogeneity according to regional manager expectations.

3.4 Firmwide Rollout

The firm was very pleased with the results of the RCT. The research team presented preliminary results from the RCT to the study firm in December 2021. Given the success of the RCT, the firm immediately rolled out paperwork reduction to the whole firm, implemented at the end of January 2022. Beyond the quantitative results of the RCT, the firm regularly receives informal feedback from workers and managers at the stores.

However, in the firmwide rollout, only the operational checklist was removed. The daily protocol was reinstated. A key reason was that feedback from workers and managers supported some value to having the daily protocol. Some workers and managers thought that having the protocol was useful for coordinating production (Alonso *et al.*, 2008). As of September 2022, i.e., after 9 months, the firm has continued not having the operational checklist.

Given the heterogeneity by regional managers, one interesting question is why didn't the firm implement the paperwork reduction in stores where regional managers expected it to work, but remove it in stores where regional managers did not expect it to work. There are two reasons against this. First, while there are sizable positive effects of paperwork reduction in stores where regional managers expected the treatment to work, it is not the

case that there are sizable negative effects of paperwork reduction. Thus, while paperwork reduction did not yield extra returns in stores where regional managers predicted it not to work, it is not the case that such reduction proved to be harmful. Second, while the firm thought it was logistically feasible to differentiate store procedures for the period of an RCT, the firm did not think that this would be feasible from a longer-run perspective. The firm often adds new stores, and would need to be surveying regional managers about whether the treatment would be effective in a new store, and regional managers would need to do this with limited information about the characteristics of the new store.

The message from the RCT and reinforced by the rollout is not that all paperwork is bad or worthless. Rather, the firm discovered that certain types of paperwork were not a good fit for the organization. The firm eliminated the paperwork that many workers regarded as annoying or demeaning. However, it kept the daily protocol, which helps coordinate production across shifts and days of the week.

4 Conclusion

In a large German bakery chain, we document that there is wide variation across tasks in the perceived value and time costs of paperwork. Randomly removing two of the most onerous paperwork tasks, including the daily checklist, we document that average sales increases by almost 3% and manager attrition substantially declines. Thus, in our setting, removing the checklist proved positive for performance.

We believe that the most important finding of our paper, however, concerns heterogeneity. In a novel methodological contribution, before the RCT and before treatments are assigned, regional managers predict in which of their stores the treatment will be effective. We show that the benefits of store sales and employee attrition are entirely concentrated in stores where regional managers predict the treatment will be effective. These predictions are only weakly explained by observable characteristics of stores. This suggests that managers may have tacit or private knowledge about which stores are most likely to benefit from a particular management practice, namely, paperwork reduction.

How might the effects we observe generalize to other context? Like all RCTs, our results are specific to our organizational context, namely, a leading firm in the German bakery chain industry. Our results suggest that the impacts of paperwork reduction and of eliminating checklists in particular are likely to be highly heterogeneous. In organizational contexts where problems observe frequently or where the costs of problems are severe, checklists may play a critical role, and eliminating checklists could be harmful. On the other hand, in contexts where having a checklist is time-consuming or is interpreted as a sign of disrespect,

we believe that eliminating checklists may be beneficial.

Note that we are not arguing that checklists are always bad in bakery chains. As seen in Figure 1, the firm uses a number of other checklists and some of them like the hygiene checklist are regarding as having a beneficial mix of time cost and value.

The RCT lasted for 9 months before paperwork reduction was rolled out firmwide. This is a long period of time compared to most management practice RCTs, and the impact on sales is apparent in the months 7-9 of the RCT.

We look forward to future RCTs examining the direct and indirect costs of monitoring. Our results also suggest that it might be useful in a variety of context to elicit subjective beliefs regarding the likely effect of an RCT on particular individual units like people or stores.

References

Aghion, Philippe, & Tirole, Jean. 1997. Formal and Real Authority in Organizations. *Journal of Political Economy*, **105**(1), 1–29.

Alan, Sule, Corekcioglu, Gozde, & Sutter, Matthias. 2023. Improving Workplace Climate in Large Corporations: A Clustered Randomized Intervention. *Quarterly Journal of Economics*, Forthcoming.

Alonso, Ricardo, Dessein, Wouter, & Matouschek, Niko. 2008. When does coordination require centralization? *American Economic Review*, **98**(1), 145–79.

Bandiera, Oriana, Best, Michael Carlos, Khan, Adnan Qadir, & Prat, Andrea. 2021. The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats. *Quarterly Journal of Economics*, **136**(4), 2195–2242.

Belot, Michele, & Schröder, Marina. 2016. The spillover effects of monitoring: A field experiment. *Management Science*, **62**(1), 37–45.

Benabou, Roland, & Tirole, Jean. 2003. Intrinsic and Extrinsic Motivation. *Review of Economic Studies*, **70**(3), 489–520.

Blader, Steven, Gartenberg, Claudine, & Prat, Andrea. 2020. The Contingent Effect of Management Practices. *Review of Economic Studies*, **87**(2), 721–749.

Bloom, Nicholas, & Van Reenen, John. 2007. Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics*, **122**(4), 1351–1408.

Bloom, Nicholas, Sadun, Raffaella, & Van Reenen, John. 2012. The Organization of Firms Across Countries. *Quarterly Journal of Economics*, **127**(4), 1663–1705.

Bloom, Nicholas, Liang, James, Roberts, John, & Ying, Zhichun Jenny. 2014. Does Working from Home Work? Evidence from a Chinese Experiment. *QJE*, **130**(1), 165–218.

Bloom, Nicholas, Brynjolfsson, Erik, Foster, Lucia, Jarmin, Ron, Patnaik, Megha, Saporta-Eksten, Itay, & Van Reenen, John. 2019. What Drives Differences in Management Practices? *American Economic Review*, **109**(5), 1648–83.

Bruhn, Miriam, & McKenzie, David. 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *AEJ: Applied*, **1**(4), 200–232.

Cai, Jing, & Wang, Shing-yi. 2022. Improving Management through Worker Evaluations: Evidence from Auto Manufacturing. *Quarterly Journal of Economics*, **Forthcoming**.

de Rochambeau, Golvine. 2020. *Monitoring and Intrinsic Motivation: Evidence from Liberia's Trucking Firms*. Mimeo, Science Po.

DellaVigna, Stefano, & Pope, Devin. 2018. Predicting Experimental Results: Who Knows What? *Journal of Political Economy*, **126**(6), 2410–2456.

DellaVigna, Stefano, Pope, Devin, & Vivalt, Eva. 2019. Predict science to improve science. *Science*, **366**(6464), 428–429.

Dickinson, David, & Villeval, Marie-Claire. 2008. Does Monitoring Decrease Work Effort?: The Complementarity Between Agency and Crowding-out Theories. *Games and Economic behavior*, **63**(1), 56–76.

Duflo, Esther, Hanna, Rema, & Ryan, Stephen. 2012. Incentives Work: Getting Teachers to Come to School. *American Economic Review*, **102**(4), 1241–78.

Ellingsen, Tore, & Johannesson, Magnus. 2007. Paying Respect. *Journal of Economic Perspectives*, **21**(4), 135–150.

Ellingsen, Tore, & Johannesson, Magnus. 2008. Pride and Prejudice: The Human Side of Incentive Theory. *American Economic Review*, **98**(3), 990–1008.

Falk, Armin, & Kosfeld, Michael. 2006. The Hidden Costs of Control. *American Economic Review*, **96**(5), 1611–1630.

Frederiksen, Anders, Kahn, Lisa B., & Lange, Fabian. 2020. Supervisors and Performance Management Systems. *Journal of Political Economy*, **128**(6), 2123–2187.

Friebel, Guido, Heinz, Matthias, Krueger, Miriam, & Zubanov, Nikolay. 2017. Team Incentives and Performance: Evidence from a Retail Chain. *American Economic Review*, **107**(8), 2168–2203.

Friebel, Guido, Heinz, Matthias, & Zubanov, Nikolay. 2022. Middle Managers, Personnel Turnover, and Performance: A Long-Term Field Experiment in a Retail Chain. *Management Science*, **68**(1), 211–229.

Friebel, Guido, Heinz, Matthias, Hoffman, Mitchell, & Zubanov, Nick. 2023. What Do Employee Referral Programs Do? Measuring the Direct and Overall Effects of a Management Practice. *Journal of Political Economy*, **forthcoming**.

Gawande, Atul. 2010. *The Checklist Manifesto*. Picador.

Haaland, Ingar, Roth, Christopher, & Wohlfart, Johannes. 2022. Designing Information Provision Experiments. *Journal of Economic Literature*, **Forthcoming**.

Halac, Marina, & Prat, Andrea. 2016. Managerial attention and worker performance. *American Economic Review*, **106**(10), 3104–32.

Hoffman, Mitchell, & Burks, Stephen V. 2020. Worker overconfidence: Field evidence and implications for employee turnover and firm profits. *Quantitative Economics*, **11**(1), 315–348.

Hoffman, Mitchell, & Tadelis, Steven. 2021. People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy*, **129**(1), 243–285.

Holmstrom, Bengt. 1979. Moral Hazard and Observability. *The Bell Journal of Economics*, 74–91.

Hubbard, Thomas N. 2000. The demand for monitoring technologies: The case of trucking. *The Quarterly Journal of Economics*, **115**(2), 533–560.

Hubbard, Thomas N. 2003. Information, Decisions, and Productivity: On-Board Computers and Capacity Utilization in Trucking. *American Economic Review*, **93**(4), 1328–1353.

Ichniowski, Casey, & Shaw, Kathryn. 1999. The Effects of Human Resource Management Systems on Economic Performance: An International Comparison of US and Japanese Plants. *Management Science*, **45**(5), 704–721.

Ichniowski, Casey, Shaw, Kathryn, & Prennushi, Giovanna. 1997. The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines. *AER*, **87**(3), 291–313.

Jackson, C. Kirabo, & Schneider, Henry S. 2015. Checklists and worker behavior: A field experiment. *American Economic Journal: Applied Economics*, **7**(4), 136–68.

Kelley, Erin M., Lane, Gregory, & Schönholzer, David. 2021. *Monitoring in Small Firms: Experimental Evidence from Kenyan Public Transit*. Mimeo, IIES.

Ko, Henry CH, Turner, Tari J, & Finnigan, Monica A. 2011. Systematic review of safety checklists for use by medical care teams in acute hospital settings-limited evidence of effectiveness. *BMC health services research*, **11**(1), 1–9.

Lazear, Edward P., Shaw, Kathryn, & Stanton, Christopher. 2015. The Value of Bosses. *Journal of Labor Economics*, **33**(4), 823–861.

Levitt, Steven D., & List, John A. 2011. Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. *American Economic Journal: Applied Economics*, **3**(1), 224–238.

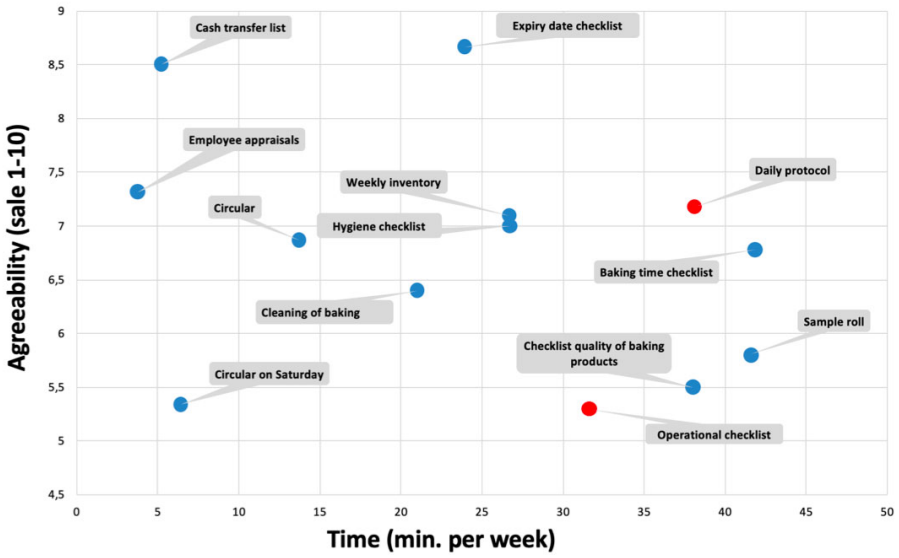
Milgrom, Paul, & Roberts, John. 1990. The Economics of Modern Manufacturing: Technology, Strategy, and Organization. *American Economic Review*, **80**(3), 511–528.

Nagin, Daniel, Rebitzer, James B., Sanders, Seth, & Taylor, Lowell J. 2002. Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment. *American Economic Review*, **92**(4), 850–873.

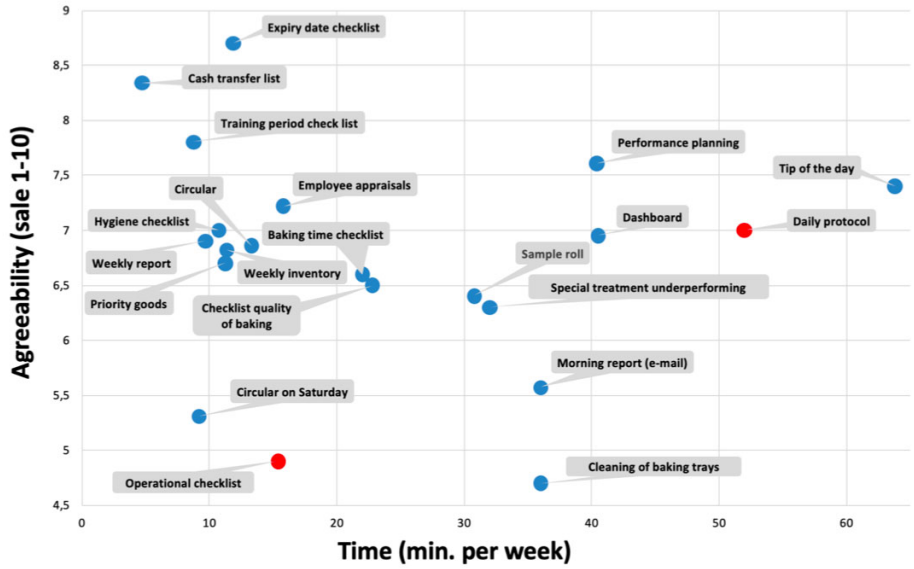
Strausz, Roland. 2006. Buried in paperwork: Excessive reporting in organizations. *Journal of Economic Behavior & Organization*, **60**(4), 460–470.

Syverson, Chad. 2011. What Determines Productivity? *Journal of Economic Literature*, **49**(2), 326–65.

Figure 1: Variation Across Tasks in Value and Time per Week



(a) Workers



(b) Managers

Notes: This figure shows answers to the below question: "The documentation duty helps (FIRM) to get better and reach company goals."

Figure 2: Operational Checklist from December 2020 (i.e., the month when the top management decided to conduct the RCT with the research team). Bolding and highlights from the original.

	Mo	Tue	Wed	Thu	Fr	Sat	Sun
1. Covid							
a) Current **covid guidelines** followed! Collecting customer contacts, serving customers: gloves, wearing face mask, keeping distance, airing out the shop	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
b) Covid hotline: PHONE NUMBERS All questions concerning covid, quarantine, sickness pay	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
2. Opportunities to increase sales							
a) Spelt products initiative phase 2 Hand over all new spelt flyers to all customers, but do **NOT** put them in the bread bag! **Please destroy old flyers** Recall: Spelt products are: LIST OF 12 DIFFERENT PRODUCTS	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
b) Bring your own cup initiative correctly implemented? For additional cups contact your regional manager	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
c) Snack of the month December Cheese-ham-cabbage → Be aware of combined offers	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
d) Please be mindful of the **appearance of the Berliner doughnut**. In a recent store visited, the sugar was partly scraped off on the side of a Berliner. Carefully touch the Berliners with a cake tong on the side; never touch a Berliner with the cake tong on the top, as sugar might be scraped off; monitor other reasons why sugar is scraped off on Berliners	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
e) **Roasted almonds correctly placed** Loosely placed on a baking tray in the cake counter, on top of 2-4 packed, not yet closed bags of almonds	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
f) **Christmas cookies** **Sufficient amount of the mini spelt almond cookies?** **→ If you do a free sample, put 4 mini spelt almond cookies in a 1 kg bag and hand it to the customers!** Sufficient amount of Christmas bags 4 kg Sufficient amount of all Christmas cookies? Follow order processes! Product assortment: - Cookie basket on top pf the counter: All types of almond cookies, coconut cookies, shortbread cookies (5 types) - Edge of the cake counter: Tree cake, gingerbread, Christmas cake - In the counter: alternating between puff cookies and shortbread cookies	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
g) Product trial Blueberry-pudding snack in LIST OF SHOPS	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
3. Organizational implementation tasks							
a) **New bonus system for wasted & returned goods** since Dec 1st Make sure to check every day If you have questions, contact your regional manager	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.
b) **Coffee bags** When making and selling coffee, please first empty **old** coffee bags before opening new ones	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.	Sign.

Figure 3: Daily Protocol from December 2020. Bolding and highlights from the original.

Date: _____ Store: _____

	Cash register number	Cash ACTUAL	Cash TARGET	Difference	Sign.	Safe bag	
						Banknote	Coins
1.		€	€	€			
2.		€	€	€			
3.		€	€	€			
4.		€	€	€			
5.		€	€	€			
6.		€	€	€			
7.		€	€	€			
8.		€	€	€			
9.		€	€	€			
10.		€	€	€			
11.		€	€	€			
12.		€	€	€			

Sales (€)		Working hours		Performance	

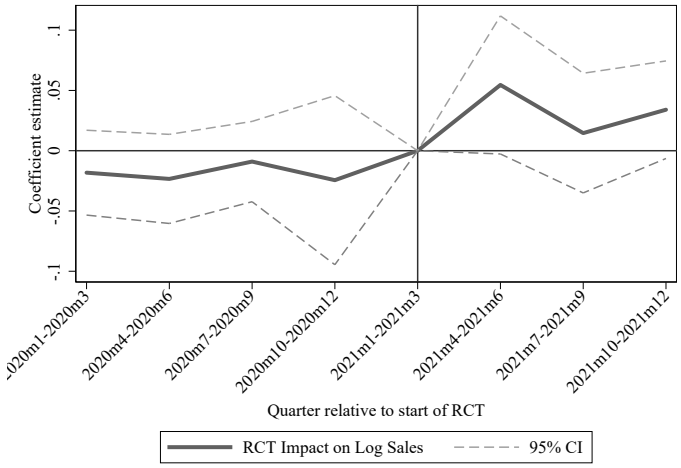
Special orders "sold out" → should we order more?

Facility or IT problems, etc.

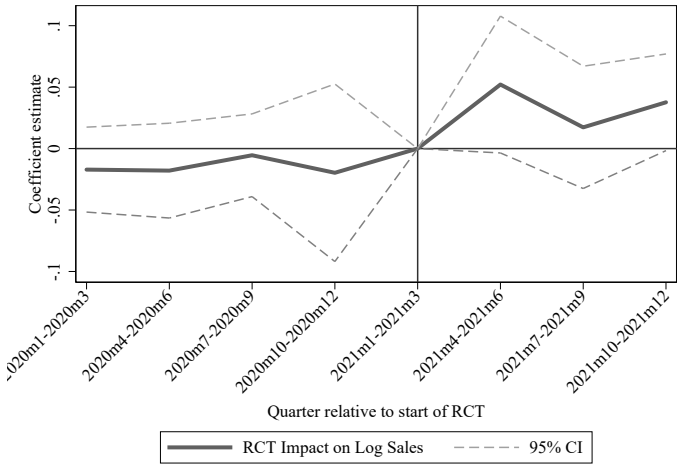
Shift changes

Additional information for the next shift

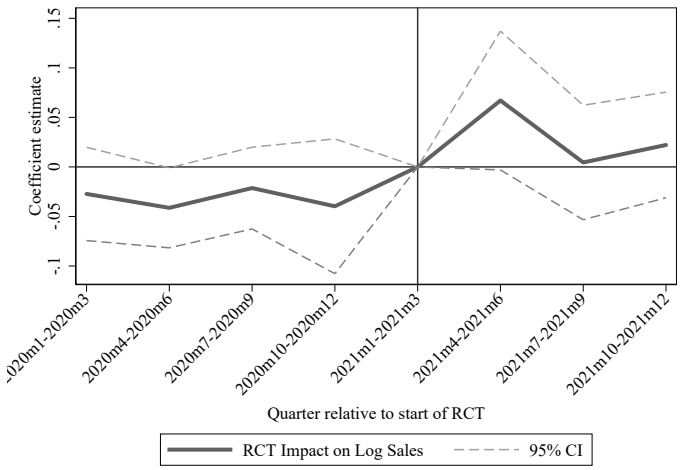
Figure 4: Event Study Impacts of the Treatment: Stores Where Treatment Expected to Have Effect



(a) Log Sales



(b) Log Busy Sales



(c) Log Slow Sales

Notes: This figure shows the event study impacts of paperwork reduction.

Table 1: Comparing Pre-Treatment Store Means across the Treatment Groups ($N = 150$ stores): Randomization Check

	Log Sales	Log Busy Sales	Log Slow Sales	Log Customers	Shrinkage	Mystery Shopping Score	Head count	Store League Ranking
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.02	-0.02	-0.03	-0.00	-0.00	-0.05	0.50	-4.98
	(0.05)	(0.05)	(0.07)	(0.05)	(0.03)	(0.08)	(0.78)	(7.24)
Constant	11.16***	10.83***	9.87***	9.85***	-2.06***	18.98***	13.30***	78.46***
	(0.04)	(0.03)	(0.05)	(0.03)	(0.02)	(0.06)	(0.50)	(4.86)
p-val	0.72	0.76	0.69	0.95	0.91	0.53	0.52	0.49

Notes: This table compares pre-RCT store-level characteristics across treatment and control stores. Robust standard errors in parentheses * significant at 10%; ** significant at 5%; *** significant at 1%

Table 2: Impact of the Treatment on Sales and Other Store Outcomes

Dep. var.:	Log Sales	Log Busy Sales	Log Slow Sales	Log Customers	Shrink -age	Mystery Shopping Score
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All Stores						
Treatment	0.028*	0.027*	0.034	0.020	0.000	-0.022
	(0.016)	(0.015)	(0.020)	(0.017)	(0.016)	(0.097)
Mean DV if Treat=0	11.17	10.86	9.832	9.740	-2.102	18.94
Observations	1,289	1,289	1,289	861	861	791
Stores	145	145	145	145	145	144
Panel B: Stores Where RCT Predicted to Work by Regional Mgrs						
Treatment	0.052**	0.051**	0.060**	0.049**	-0.034	0.043
	(0.021)	(0.021)	(0.023)	(0.023)	(0.023)	(0.126)
Mean DV if Treat=0	11.09	10.77	9.754	9.660	-2.068	19.01
Observations	670	670	670	447	447	416
Stores	76	76	76	76	76	75
Panel C: Stores Where RCT Predicted Not to Work by Regional Mgrs						
Treatment	-0.002	-0.001	0.010	-0.014	0.036	-0.091
	(0.021)	(0.020)	(0.034)	(0.023)	(0.022)	(0.153)
Mean DV if Treat=0	11.27	10.96	9.922	9.834	-2.141	18.85
Observations	619	619	619	414	414	375
Stores	69	69	69	69	69	69
p-Val on diff between Panels B and C	0.09	0.09	0.21	0.06	0.03	0.49

Notes: An observation is a store-month during the RCT. Standard errors clustered at the store level are in parentheses. Each regression controls for the mean of the dependent variable in the pre-period and year-month fixed effects. The p-value in the final row is calculated using an interactions. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: Impact of the Treatment on Employee Attrition

Sample:	All	Qualified Workers	Mini-jobbers	Mgrs
	(1)	(2)	(3)	(4)
Panel A: All Stores				
Treatment	0.23	-0.40	0.82	-1.15*
	(0.24)	(0.31)	(0.60)	(0.60)
Mean DV if Treat=0	2.695	1.810	4.307	1.669
Observations	16,326	6,917	4,565	1,347
Workers	2310	879	721	161
Panel B: Stores Where RCT Predicted to Work				
Treatment	0.05	-0.51	0.85	-2.09**
	(0.34)	(0.48)	(0.90)	(0.92)
Mean DV if Treat=0	2.613	1.915	4.163	2.374
Observations	7,990	3,541	2,108	748
Workers	1138	461	328	92
Panel C: Stores Where RCT Not Predicted to Work				
Treatment	0.34	-0.33	0.48	0.46
	(0.37)	(0.43)	(0.89)	(0.85)
Mean DV if Treat=0	2.777	1.682	4.437	0.763
Observations	8,336	3,376	2,457	599
Workers	1241	447	400	73
p-Val on diff between Panels B and C	0.62	0.79	0.63	0.04

Notes: An observation is a worker-month during the RCT. Standard errors clustered at the store level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Impact of the Treatment on Sales: Heterogeneity by Team Size

	(1)	(2)	(3)
Sample	Small teams	Big teams	All
Treatment	0.083**	0.012	0.081**
	(0.037)	(0.017)	(0.038)
Big team at firm			0.058**
			(0.028)
Treatment X Big Team			-0.070*
			(0.042)
Observations	305	984	1,289
Mean DV if Treat=0	10.82	11.27	11.17
Stores	35	110	145

Notes: An observation is a store-month during the RCT. Standard errors clustered at the store level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Responses from the Regional Manager Survey: Explanations for Why the Treatment Will Work

Explanation	Share
A Utility Explanation, Such as People Like Not Having Paperwork or Feeling Less Stressed About Paperwork	32%
A Problem Explanation Such as Team Having Good Communication or Not Need Any Bureaucracy or Not Experiencing Problems Communication	66%
Regional Managers Will Invest More Time if Store if it is Treated Such as Visiting or Calling More	0%
Treatment Stores are Likely to Experience Outside Shocks to Performance During the RCT	0%

Notes: These data are from the pre-RCT regional manager prediction survey. The numbers are based on examining the free text responses of regional managers.

Table 6: Heterogeneity in Sales Effects Based on Time Spent on the Daily Protocol

	(1)
Treatment	0.032
	(0.031)
Treatment X Time spent on daily protocol	-0.000
	(0.001)
Time spent by store on daily protocol, overall	0.000
	(0.001)
Observations	1,221

Notes: An observation is a store-month during the RCT. Standard errors clustered at the store level are in parentheses. Each regression controls for the mean of the dependent variable in the pre-period and year-month fixed effects. A big team means more than 10 workers at the store.* significant at 10%; ** significant at 5%; *** significant at 1%

Web Appendix, "Is This Really Kneaded? Ask the Manager! A Large-scale Trial on the Effects of Paperwork Reduction", by Friebel, Heinz, Hoffman, Kretschmer, and Zubanov

Appendix A contains additional figures and tables. Appendix B is the Data Appendix. Appendix C provides materials used by the firm in the RCT.

Appendix A Appendix Figures and Tables

Figure A1: Picture of a Sample Bakery



Table A1: Regional Manager Predictions, Part 1

Yes	Prediction
1	Would be very happy about less bureaucracy, less work as a result, do not like to work with notes and strict rules, will work
0	Both: Some employees are happy about fewer guidelines, others need strict rules
0	Will have a positive impact on employee satisfaction; but: poor communication of initiative by branch management expected, might have negative impact on s
1	Great, well-coordinated team in the branch, everything fits in the branch, would appreciate less bureaucracy
0	Both: employees will be happy, but individual employees need more restrictions
1	Well-coordinated team, has been working together for a long time, very good communication within the team, would be glad, no negative effects; will work!
0	Negative effects, as the team is still very fresh, new management in place, processes not yet internalised; Negative sales
0	Both. Employees will be glad, mixed team with some old and many young employees.
1	Would perhaps miss the list; but: no negative consequences in the branch; on the contrary: positive impact!
0	Will be glad; but: implementation of processes not secure, chaotic branch; internal evaluations (e.g. strawberries on a cake) mostly negative. Might be chaotic
1	Would implement this very well, would also get along well without paper and clear structure; employee satisfaction will increase
0	Many new staff members, branch is a bit chaotic, need structure and guidance, want guidance
1	Get along without bureaucracy; would feel more comfortable if there was less pressure because of less bureaucracy. Will work
1	Get along without bureaucracy, nothing would change in the operational processes without bureaucracy, staff already understood important things
0	Mixed picture; have too high returns on bakery products, returns will get worse. Unclear how it will work
1	Get along without bureaucracy, nothing would change. Therefore, will work
0	Need structure, will not work without it, otherwise the branch will sink into chaos and lose focus
0	Need structure, haven't been around long, bureaucracy is important support, returns on bakery products
0	Need structure and bureaucracy, otherwise staff will have problems
1	Yes, will work
1	Yes will work
1	Yes, will work
0	No, will not work
1	Yes, will work
0	No, does not work
0	No, does not work
1	Yes, will work. Clear yes
0	No, does not work. No way
0	No, does not work. No way
1	Will work. Good and organized branch management; very conscientious and tidy. Implementation will work
0	Need assistance. Complicated without lists, young branch management, young team needs guidance
0	Undecided. Maintain documentation obligations, as other structure is difficult to implement; old branch management, which wants to maintain habits
1	Branch team does not need lists. Committed, thoughtful and conscientious
0	Branch desperately needs structure which is provided by bureaucracy; organized branch management , bad team. Will not work without lists
0	Good leadership, bad team. Would work partially
0	Would be good if lists remained. Recent change of management. Large branch
1	Would work. Complete Confidence in the team
1	No documentation requirements needed. Good team. Good branch
1	No documentation requirements needed. Good team and branch management. Well organized
0	Will not work - team is still finding itself; guidance and structure needed; possible problems if list isn't there anymore. If there's a mystery shopping visit and
1	Does not work in this branch as good as in branch ...., but will work as well; maybe some structure needed, also autonomous possible. Will work
1	Similar to branch in ....; team will be glad; actually need list to get routine, would also work out without list
1	Will work out without any requirements, team is confident in their performance, happy if there are no lists
1	Like in branch .... Team will manage it, but need to stay focused. Problem: When there is a Mystery Shopping visit and expectations are not met, there will
1	Team does not need lists. Can manage without lists. Strength in implementing processes.
1	No lists needed, works out without lists. However, when the branch manager is not on duty, they sometimes not meet expectations
0	List needed for orientation. Does not work without it.
1	Definitely do not need lists, implement everything either way
1	Do well without a list
1	In general: will work out
1	Will work out.
1	Could do well without lists, would like to keep daily log
0	Focus branch; cannot work without clear guidelines, may result in chaos
1	No bureaucracy needed, even if daily log is important from time to time
0	Focus branch; cannot work without clear guidelines, may result in chaos
0	Cannot work without it, cash differences
1	Can do without it, branch runs great
1	Can do without documentation requirements, runs great, but still relatively new branch management
0	Cannot do without it, big cash and branch differences and problems with sales; even if employees would like to pass on restrictions
1	Will work out without restrictions
1	Will work out without restrictions
1	Will work out without restrictions
1	Will work out without restrictions
1	Will work out without restrictions
0	Will work out without restrictions
1	Will work out without restrictions
1	Will work out without restrictions
0	Structure needed. Won't work without it
1	Will work out without restrictions
0	Staff will be glad; procedures are sometimes problematic, often not implemented, therefore bureaucracy and structure needed
0	Branch management wants to maintain bureaucracy; but it could work as well. Unclear if it works out
0	?
0	Branch wants to keep bureaucracy, unclear if it works out
0	Branch wants to maintain bureaucracy, clear structures important for training and coaching. Unclear what happens
0	Branch wants to maintain bureaucracy, important for training and coaching; mixed effects
0	Branch wants to maintain bureaucracy, important for training and coaching, has to deal with branch differences and leadership
0	Branch wants to maintain bureaucracy, important for training and coaching, has to deal with branch differences and leadership
0	Sometimes help needed, large branch, operationally strong, so could also work out
1	Can be left out, very strong branch management, trains staff very well
1	Can be left out, small branch, few employees, can also be trained in person

Notes: This table gives the first table of regional manager predictions.

Table A2: Regional Manager Predictions, Part 2

Yes	Prediction
0	New branch management old established team, need guidance, but could work out in the medium term
0	No good branch management, not good at training staff, clear guidance and lists are important
1	Works out without, small branch, staff are well trained and guided by branch management
0	Do not leave out, big team, difficult cases,information does not go down well
1	Training on important processes is also possible this way, control can be omitted, will work out
0	New branch management, lists are needed
0	New branch management, lists are needed, but branch management is probably good, best case: keep first, leave out later
1	Independent branch, will work out without lists, employee satisfaction will improve
0	Downtown branch, no positive or negative developments on sales or performance, high employee satisfaction anyway
1	Similar to other well running branches, team will be glad if lists are gone, no change in sales (maybe better sales), time is saved, no change in other numbers
0	Similar to other well running branches, team will be glad if lists are gone, no change in sales, time is saved, no change in other numbers
1	If operational list is gone, it's good for the team, it will work
1	Always enjoyed making lists and bureaucracy, but will also work out well without restrictions
0	Always enjoyed bureaucracy. Old employees and therefore difficulties without it
1	Team will be glad when operational list is gone. No problems expected. Will work out!
0	Rather neutral. Mixed effects. No operational list is good, more time for employees
0	Will not be received well,. Daily protocol and operational lists are popular; employees like bureaucracy
0	Like bureaucracy, will find another way, will neither be happy nor sad; neutral effects
0	Bureaucracy needed
1	Will work out without
1	Will work out without
0	Documentation requirements are needed
1	Could live without bureaucracy, very communicative branch management
0	Daily protocol needed, operational list not necessarily. Therefore mixed effects
0	Bureaucracy needed, will not work out without
1	Strong branch management, high sales, employee satisfaction 50/50, branch management will not take omitting lists seriously, because there are so many other
1	Strong branch management, been there for a long time, high employee satisfaction, it will work out very well without documentation requirements
0	Currently closed, strong branch management, employee satisfaction high and will improve
1	Small branch, just getting started, new branch management, will accept bureaucracy reduction and implement successfully. It's a chance!
0	Very strong branch management, employee satisfaction will not change. Large branch. But: operational implementation will work partially , no big problems
1	Strong branch management, open to everything, high employee satisfaction, omitting lists will be successful
0	Small branch, will be received positively, new branch management, mixed effects
0	Very strong branch management, employees been there for many year. Effects unclear
0	Will meet with resistance, will not accept anything new, will only reluctantly, if at all, let themselves be dragged into it, branch management communicates th
1	Strong branch management, open to everything and can implement everything well, already been there a few years
1	Employee satisfaction will improve with less bureaucracy, strong branch management, will work out
1	Interested branch management, will be happy about it, positive response, higher employee satisfaction; Omitting will work out
1	Top motivated branch management, positive response, branch management takes on many tasks itself, less bureaucracy will be supportive
1	Focal point branch, motivated branch management; branch management takes over a lot of bureaucracy from staff; employee satisfaction will not improve nece
0	Critical branch, employee satisfaction will not get better, does not work out
1	Mini branch, hardly any bureaucracy, will work out
1	Mini branch, hardly any bureaucracy only 3 employees will be happy when there is less bureaucracy
1	Branch management will be happy that lists/ bureaucracy are gone, but then say: does not help much; employee satisfaction will not increase, but it will work
1	Highly motivated branch team, very communicative, maybe no increase in sales or staff satisfaction , because branch is already productive, will work without l
0	Old branch management, if it is up to them they will continue to run lists; no change in sales, independent from restrictions - branch will be ok
1	Great branch management, will work hard on it and implement it well, will analyze whether it is successful. Will work. Positive influence; employees very satis
0	Employees are dissatisfied with the situation in the branch, there are grumblings, feeling relieved because of less bureaucracy could help, unclear what happens
1	Will work, good branch and well organized branch management
0	Problem team, a bit chaotic. Won't work without guidelines and clear guidelines
1	Could probably work, well organized branch management and team
1	Will work, but: branch management is very bureaucratic
1	Branch manager retiring soon. Could work out- well-functioning team; unclear if open to changes, but it will work in general
1	Could work, or rather: Will work!!
0	No, will not work
1	Will work. But team needs to know why
0	At the moment, no. Will not work
1	Yes, we implement well, but want to understand why. But: If explanation makes sense (which may be the case), it will work
1	Bureaucracy costs time; more time has a positive effect on satisfaction; will work out
0	Older employees, very bureaucratic, keep handwritten lists, love bureaucracy, unclear
1	Less bureaucracy saves time; more time = positive for employee satisfaction, young team, easy-going
1	Less bureaucracy saves time; more time = positive for satisfaction, young team, more relaxed and more free time
0	Structures and control needed
0	Will improve the general mood, are often overwhelmed with bureaucracy; employee satisfaction and sales will not improve
0	Undecided
0	Branch management over 20 years in, undecided
0	Less bureaucracy will improve the general mood; but: employee satisfaction and sale will not improve. Unclear what happens
0	Undecided
0	Very communicative, new branch, needs bureaucracy

Notes: This table gives the second table of regional manager predictions.

Appendix B Data Appendix

Store league performance ranking. Inspired by the German Bundesliga, the firm uses different measures of performance to provide an overall scores to stores at the firms, and the stores are then ranked. The goal is to account for differences in possible sales and profitability.

Appendix C Materials Used in the RCT and Firmwide Rollout

C.1 Wording Used for the Regional Manager Predictions

I presented the pilot project in a regional manager meeting in Feb 2021. I received the following feedback about the pilot project from the regional managers:

"In some shops, less documentation duties will work well in the daily business operations and will probably have a positive effect on store performance indicators. In other shops the reduction will have negative effect on the daily business and will probably have a negative impact on store performance indicators."

We as researchers are interested in your predictions!

Now I will ask you to make predictions for all of your shops (independent whether the shop will indeed be a pilot shop or not).

I have now a list of your shops (in front of me)

What do you think: If the shop XYZ indeed was a pilot shop: How well would the daily business work ("function") in the shop with less documentation duties?