Making the Right Call: The Heterogeneous Effects of Individual Performance Pay on Productivity*

Marco Clemens† Jan Sauermann‡

August 11, 2023

Abstract:

Performance pay has been shown to have important implications for worker and firm productivity. In this study, we apply a dynamic difference-in-differences estimator to estimate the causal effects of the introduction of a generous bonus pay program with salient performance thresholds on incentivized and non-incentivized performance outcomes in a call center. On average, the performance pay program did not affect workers' performance. We show, however, that this result conceals an underlying heterogeneity in the response to performance pay: High-skilled workers are more likely to meet the performance targets, while low-skilled workers are less likely to do so and might even perform worse in the non-incentivized outcome. The findings can be rationalized with the idea that the costs of effort differ for individual workers. We also explore whether agents alter their overtime hours and find a negative effect, possibly avoiding lower call quality through longer working hours.

Keywords: performance pay, incentives, productivity, skills, panel data
JEL-codes: M52, J33, C23

---

*We thank Lazlo Goerke, and seminar and conference participants at the RES PhD Conference 2023 (Glasgow), IAAE 2023 (Oslo), WLE 2023 (Trier) and two anonymous reviewers of the HWS 2023 (Berlin) for valuable comments and feedback. Declarations of interests: none.

† *Corresponding author*. Institute for Labor Law and Industrial Relations in the European Union (IAAEU) and Trier University; Behringstraße 21, Trier; e-mail: clemens@iaaeu.de

‡Institute for Evaluation of Labor Market and Education Policy (IFAU), Uppsala Center for Labor Studies (UCLS), and the Institute of Labor Economics (IZA); jan.sauermann@ifau.uu.se

1 Introduction

Performance-related pay remains an important part of workers' pay: in the US, 29% of all employees with a college degree receive pay that at least partly depends on individual, team or firm performance (Maestas et al., 2017). The increasing availability of performance measures and the increasing use of the option to work from home will likely make performance-related pay more prevalent (Eurofund, 2020; Barrero et al., 2023). At the same time, performance pay has important implications for human capital acquisition (see, e.g., Camargo et al., 2022; Taylor, 2022) and earnings inequality (see, e.g., Lemieux et al., 2009; Bryan and Bryson, 2016). The literature on performance pay based on worker-level performance has shown that introducing performance-related pay often positively affects worker performance, firm productivity, and worker turnover (see, e.g., Lazear, 2000; Eriksson and Villeval, 2008; Lavy, 2009; Gielen et al., 2010; Dohmen and Falk, 2011; O'Halloran, 2012; Manthei et al., 2022), but also that the effects differ substantially between studies (Havranek et al., 2022). While heterogeneity in treatment effects can arise due to various factors, such as the size of the monetary incentive (Gneezy and Rustichini, 2000), individual risk-preferences (Cadsby et al., 2007; Grund and Sliwka, 2010), or whether the outcome is qualitative or quantitative, performance-related pay can even generate heterogeneous impacts *within* workplaces if managers are incentivized (Bandiera et al., 2007) or if workers differ by skill level (Kowalski, 2019).

In this study, we analyze the introduction of performance bonuses in the call center of a multi-national telephone company located in the Netherlands to study if and how individual performance bonuses affect worker performance. We employ dynamic difference-in-difference methods (Sun and Abraham, 2021) to compare workers eligible for performance pay with those ineligible. The aim of this paper is threefold: First, we show that the skill distribution of workers matters for the effectiveness of performance pay programs. Compared to low-skilled individuals, high-skilled individuals have a lower cost of effort to reach performance thresholds. Second, we estimate the effects of the performance pay program on the incentivized *quality* measure of performance as well as spillover effects on the non-incentivized *quantity* outcome. Third, we also estimate effects of the performance pay program on working hours. Workers can have an incentive to work extra hours in order to achieve bonus threshold levels.

We aim at contributing to the literature in several ways. First, we contribute to the literature documenting effects of performance pay programs on performance. Individual performance incentives have been shown to increase effort, but also to affect sorting of workers to different types of jobs (Lazear, 2000; Dohmen and Falk, 2011). A large number of empirical studies using personnel data for different firms and different tasks provided empirical support for this result. This includes studies using data of windshield installers (Lazear, 2000), workers in tree planting and tree thinning (Shearer, 2004; Shi, 2010), university professors (Heywood et al., 2011), and medical typing workers (Unger et al., 2020), among others. Only a few studies have examined heterogeneous effects of performance pay with respect to the underlying distribution of workers' skills. Franceschelli et al. (2010) use data from a textile company to classify workers as low-skilled if they want a base-wage scheme and high-skilled if they seek a bonus-related scheme. The results show that there is no significant difference in performance between high- and low-skilled workers. Conversely, conducting a field experiment with warehouse workers in the U.S., Kowalski (2019) finds that the strongest (positive) effects are found for low performers, followed by high performers, and the weakest effects are found for middle performers.[1] In our study, we find that the pre-treatment skill distribution of workers matters for the introduction and effectiveness of performance incentives. While the performance pay program does not affect employee performance on average, high-skilled workers respond more positively to the introduction of performance pay. The underlying intuition behind this result is that if the expected effort costs to meet the incentive thresholds are disproportionately high compared to the bonus awarded, workers may refrain from adjusting their performance. This suggests that it is important to calibrate performance thresholds appropriately.

[1] Furthermore, Manthei et al. (2021) have found that prior learning and tenure are related to the effectiveness of performance pay. Bandiera et al. (2007) show that manager incentives can create similar heterogeneity among workers if managers choose workers with higher ability over those with low ability. A similar observation has been made by Azmat and Iriberri (2010) in an educational context. With the implementation of feedback on relative performance under a piece-rate scheme, the authors show that the strongest increases in performance occur at the tails of the ability distribution.

Second, this study contributes to the literature by analyzing the effect on both incentivized measures as well as spillover effects on non-incentivized performance measures. If workers' tasks consist of several dimensions of task performance, one of which is incentivized while the others are not, workers may simply aim at performing well on the performance outcome that is rewarded (Holmstrom and Milgrom, 1991). Only a few studies analyze the effect on alternative performance outcomes not subject to the incentive introduced (Asch, 1990; Shi, 2010; Al-Ubaydli et al., 2015; Hong et al., 2018). Our study is able to analyze whether an increase in the incentivized service quality comes at the cost of a lower work speed. We find only weak evidence that this is the case for high-skilled agents. The results, however, show that, in addition to not increasing performance in the incentivized outcome, agents at the lower end of the skill distribution perform worse in the non-incentivized outcome, suggesting that they shift effort towards the incentivized outcome. We thereby also contribute to the literature by studying the effects of introducing performance pay on *quality*-related measures of performance. Compared to quantitative performance outcomes, it is typically more difficult for firms to monitor the quality of workers' output. Prior studies have used information on the educational level of conscripts (Asch, 1990), on defect rates (Kato and Shu, 2008), and quality of planting trees (Shi, 2010). If quality (or a creative task) is directly incentivized, recent evidence mainly comes from the lab with positive effects (Bradler et al., 2019), but also negative effects (Kleine, 2021) in which financial incentives even harm creative breakthroughs. The call center analyzed in this study applies a monitoring system for service quality, which is based on randomly called-back customers evaluating the quality of their call with an individual agent.

Third, we contribute to studies analyzing the relationship between performance pay and working hours. In an educational context, Angrist and Lavy (2009) demonstrate in a randomized field trial providing a cash incentive for passing an exam, can lead to extra time being devoted to exam preparation. Recent studies in a job context have strengthened those results showing that working hours may increase in response to performance-pay (see, e.g., Artz and Heywood, 2023; DeVaro, 2022; Green and Heywood, 2023). Our findings show that this channel may not apply when quality is rewarded, as workers reduced excess hours in response to the bonus to meet the performance target.

Section 2 introduces the company and incentive plan of interest, while Section 3 presents the data. Section 4 covers the methodology, Section 5 presents the results, and Section 6 provides the conclusion

2 The firm and its incentive schemes

In this section, we shortly describe the firm and the structure of their performance incentives before and after the change analyzed in this study.

2.1 The firm and workers' tasks

The call center is part of a multi-national telephone company located in the Netherlands.[2] It is an in-house service center that handles the inbound calls of current and prospective customers and is organized into different departments which cover different customer and call types. Customers can call in when they have questions or would like to report technical, administrative problems and complaints. An automated routing system connects customers to available agents.

The main task of call agents is to handle inbound customer calls. In addition to answering calls and talking to customers, agents are also required to access and enter information in the customer database. Agents are not involved in other tasks, such as written customer correspondence. Agents are assigned to team leaders whose main task is supervising the agents and monitoring their calls, i.e., there is no team specialization. Team leaders report to and are evaluated by their respective department manager.

2.2 Incentives in the firm

Although an agent's performance is measured along several dimensions for any given time interval, the firm avoided to introduce performance incentives with explicitly defined

[2]Data from the same firm are also used in, e.g., De Grip et al. (2016) and Sauermann (2023). In contrast to these papers, our paper uses a different and so far entirely unused sample that covers the introduction of performance bonuses and more departments.

performance targets for several years.[3] Instead of explicit performance incentives, the call center relied on annual appraisal interviews with the respective team leader, typically held in April and May, which resulted in a grade from one (lowest) to five (highest). Although there were no specific guidelines on the weighting of different (measurable) performance outcomes or explicitly stated goals, the highest weight was typically put on average handling time and measures related to customer satisfaction. According to management, team leaders were also supposed to evaluate an agent's behavior towards peers, team leaders, and managers during the previous year. While team managers were asked to reach a bell-curve distribution of performance ratings in their teams, the rating was at the discretion of the team leader and not (directly) dependent on an agent's observable performance outcomes.[4] This grade was then used as a multiplier for the reference wage increase and the reference bonus level. If management set the reference wage increase at 4%, a grade 1 agent would receive no wage increase, grade 3 agents would receive 4%, and agents with the highest grade (five) would receive 6% (150% of the reference wage increase). The annual bonus was calculated in the same way and could be up to a maximum of 8% of the annual wage. There is no additional seniority-related wage increase employed at the call center.

In April 2011, the annual bonus was replaced by a new, short-term incentive pay with explicitly defined performance thresholds that were salient to the agents.[5] Under the new incentive scheme, management set performance thresholds in advance of a quarter, evaluated agents' performance, and paid out a bonus of up to 12% of an agent's wage during that quarter after the end of a quarter. Compared with the incentive structure before the introduction in April 2011, this meant a drastic change to a bonus that was much more short-term and explicit. The idea behind the introduction of a bonus explicitly related to service quality is that the quality of services provided to customers has become a unique competitive advantage in the mobile communications market.

[3] According to the call center's management, the main reason for the non-utilization of explicit incentives based on observable performance outcomes was the position of the workers' council.

[4] Additional data on the performance ratings show that 54% of agents received a three, 30% received a two and 14% a four, and only 2% received a one or five.

[5] Annual appraisal interviews were still used to determine the annual wage increase.

The new performance pay is based on a measure of service quality gathered from customer satisfaction surveys of clients who were randomly called back, calculated for three months at a time (one quarter). There are five bonus levels $B_1, \ldots, B_5$, which correspond to bonuses of 0, 4.8, 8, 10, and 12% of the wages earned during the quarter, respectively (see Figure A1). If an agent's average performance during the quarter $\overline{y}$ does not exceed the lowest threshold y_1 ($\overline{y} \leq y_1$), the agent receives no bonus ($B_1 = 0\%$). Agents who outperform the highest threshold ($y_{J-1} \leq \overline{y}$) receive the highest bonus, B_J.[6]

Before the quarter, thresholds are explicitly communicated with agents. Throughout the quarter, agents are given feedback about their individual performance about once a week. This implies that the structure of the bonus pay, performance thresholds and individual performance are made salient to the agents.

While a large number of agents received performance bonuses based on their individual performance, some agents were exempted. First, agents with short tenure and agents with temporary help agent contracts were exempted. Second, agents whose quarterly performance is based on fewer than 60 evaluations were exempted. These agents received bonuses based on their department's average performance.

3 Data

The data used in this study provide monthly information on the performance outcomes of agents nine months before and after the introduction of the bonus related to service quality that took place in April 2011. All performance measures are available at the individual agent level. We will introduce the service *quality*, work speed and skill measures and offer the final estimation sample with descriptive evidence in the subsections that follow.

[6]In accordance with the bonus payments, the performance thresholds on which the bonuses are based are not equally distributed. The distance between the lowest threshold (threshold 1) and the second lowest threshold (threshold 2) on the service quality index is 0.05 units of service quality, defined on a scale of zero to one. However, the distance between thresholds 2 and 3 and that between thresholds 3 and 4 is only 0.025. Average performance differs by department. The target size is therefore adjusted accordingly. The absolute distance between the target thresholds is the same for each department.

3.1 Outcome Variables

Call centers typically have several performance measures covering different dimensions of an agents' productivity. In our data, performance can be measured as quality-related (*service quality*), and quantity-related (*work speed*). Both dimensions of performance, service quality and work speed are important to the firm because they affect customer loyalty and the total costs of the calls (wages), respectively.

Service Quality The newly-introduced bonus pay in the call center is based on a measure generated from a customer satisfaction survey among a randomly chosen population of customers. Among other questions, customers were asked to answer on a scale from zero ('very unlikely') to 10 ('very likely') how likely they would recommend the mobile operator to family and friends, based on the previous call.[7] From all answers to this question in period t, the management then calculates the net-promoter score (NPS) as the percentage point difference between the share of customers rating the agent as nine or 10 (high service quality) and those rating the agent as six or lower (low service quality):

$$y_{i,t}^{QUAL} = \frac{N_{it,9-10} - N_{it,0-6}}{N_{it,0-10}} \tag{1}$$

where N is the number of evaluations and the subscript denotes the grade given by the customer for agent i in month t. The NPS measure can then be used as a performance metric for an agent's service quality. To ease interpretation, we re-scale the NPS measure to be defined from 0 (lowest service quality) to 1 (highest service quality) throughout the paper.

Before interacting with an agent, customers were asked whether they would be willing to participate in a customer satisfaction survey. It is important to note that the agent did not know whether the customer agreed to participate in the survey or not. Neither agents nor managers can affect the selection of customers who rate agent performance and thus cannot influence this outcome measure by selection into the survey. Shortly

[7] The exact question was 'Based on this contact, how likely are you to recommend [the firm] to your family and your friends?'. The NPS is defined such that it is correlated to customer loyalty (see, e.g., Keiningham et al., 2007).

after the end of the customer call, a random subset of customers was automatically called back. An interactive voice response system then guided the customers through the survey. As service quality is derived from the customer survey, the number of evaluations relative to the actual number of calls made is relatively low. As a consequence, service quality measures are often based on only a low number of actual evaluations and thus have considerably more variation than measures that are based on all calls, such as the measure of work speed. For this reason, all estimations control for the number of customer evaluations (see Section 4). Another potential concern about using the NPS as a performance outcome is that it may be biased because of customer non-response. Agents providing low-quality (high-quality) service would then be characterized by service quality that is higher (lower) than their actual service quality, because the sample of evaluated calls is less representative of their calls than for agents providing higher (lower) quality. While this may be a potential concern about the validity of service quality as a proxy for the provided quality, there is no reason to assume that customer non-response changed with the introduction of individual performance pay.

Work Speed To estimate the effect on work speed, which was not targeted by the new incentive scheme, we use a measure that is based on the average length of calls to measure performance. Hence, we define the quantitative service outcomes as y_{it}^{QUAN}, indicating the average handling time in seconds. Similar performance measures have been used by other studies using call center data (see, e.g. Liu and Batt, 2007; De Grip and Sauermann, 2012; Battiston et al., 2021).

The variable provides a clear and objective measure of quantitative performance that is available for each agent i and all calendar months t. It measures the average time an agent spends talking to a customer and logging the information on the call in the customer database. Shorter average handling times are associated with higher performance because short calls are less costly to the firm.

3.2 Workers' Skill

One aim of this paper is to estimate the heterogeneous impacts of individual performance incentives on performance. We argue that for high-skilled workers, the effort costs of reaching pre-defined performance thresholds are lower than for low-skilled workers.

A direct measure of workers' skill is not available in the data. Instead, we use residualized service quality in the pre-treatment period, i.e. from October 2010 to March 2011.[8] For all observations during this period, we estimate the following regression equation:

$$y_{i,t}^{QUAL} = \alpha_1 + \alpha_2 X'_{it} + \varepsilon_{i,t} \tag{2}$$

with X'_{it} comprising a vector of control variables including contractual working hours, tenure, gender, age, department dummies, the number of evaluated calls, a dummy for being a temporary help agent and month-fixed effects.

For each agent-month observation, we extract the residuals $\widehat{\varepsilon}_{i,t}$, indicating quality differences that are unexplained by the aforementioned control variables. We then define agents as high-skilled if $\widehat{\varepsilon}_i$ is greater than the median $\widehat{\varepsilon}_i$, and define agents as low-skilled otherwise.[9]

3.3 Estimation Sample

As the measure of skill based on Equation (2) relies on agents being observed between October 2010 and March 2011, our sample excludes all those agents that left the firm before October 2010 and those that entered the firm only in and after November 2011.[10] After making those adjustments and removing missing values for both outcome and control variables (see Section 4), the total number of agents in the estimation sample is 388 with 3,517 agent-month observations.

[8] See Gompers et al. (2010) and Weidmann and Deming (2021) for similar approaches.

[9] In Section 6, we provide additional robustness checks regarding our skill definition. First, we extend that specification using a tertile sample split instead. Second, we also provide an alternative approach using the full pre-bonus period.

[10] Potential issues with the sample restriction are addressed in Section 6.

4 Methodology

4.1 Treatment and Control Groups

In the estimation sample, we define treatment and control groups as follows: We assign agents to the treatment group if they received the individual bonus for their performance in a given month. This specification includes cases where agents switched from temporary help agents to fixed employment contracts or have exceeded the six-month threshold in order to be eligible for treatment within a given quarter or become eligible only in later quarters after the introduction.

Figure 1 displays agents by their treatment status over time. The figure reveals that most agents are observed between October 2010 and March 2011, as the skill measure is based on those individuals. Additionally, in the first treatment quarter, only approx. 30% of agents received treatment, which increases to around 70% in the last sample month, confirming that some agents become eligible for the bonus only in later periods.

Figure 1: Distribution of Treatment and Control Groups over Time

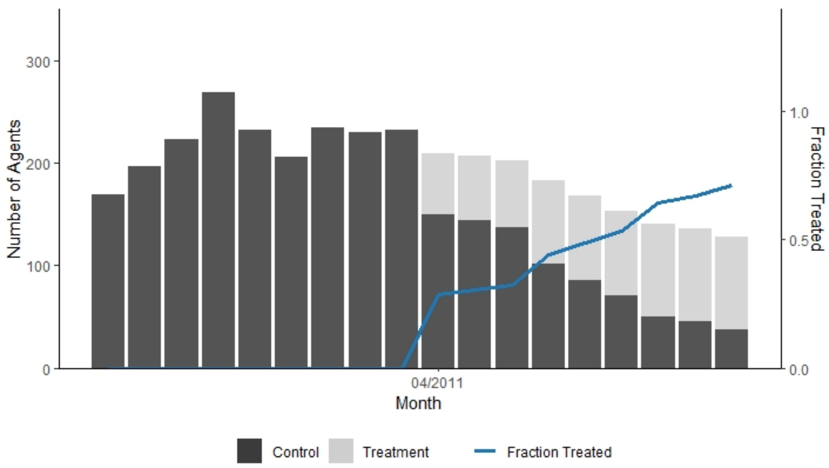


Figure notes: The figures present the distribution of treatment status over the sample period October 2010 and December 2011. Individuals are defined as treated if they report a bonus payment in a given month.

4.2 Descriptive Evidence

Table 1 shows differences in characteristics between eventually treated and never-treated individuals in the last observed month before the bonus system was introduced for the first time (March 2011). Columns (1) and (2) show the mean and standard deviation for the control group, while Columns (3) and (4) indicate the same for the treatment group. The difference in means and its statistical significance are shown in Columns (5) and (6).

Table 1: Treatment and Control Groups - Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)	(6)
	Never Treated		Eventually Treated			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff.	p-value
Worker characteristics						
Temporary Agent	0.61	0.49	0.50	0.50	-0.11	0.07
Female	0.41	0.49	0.48	0.50	0.07	0.23
Age	30.48	9.56	32.18	10.64	1.70	0.16
Tenure	2.65	3.90	3.22	4.11	0.57	0.22
Hours planned	26.84	8.84	31.72	7.71	4.87	<0.01
Evaluated calls	3.43	2.26	5.56	2.61	2.13	<0.01
Outcome variables						
$y_{i,t}^{QUAL}$	0.40	0.22	0.40	0.12	0.01	0.73
$y_{i,t}^{QUAN}$	348.37	143.96	325.05	55.84	-23.32	0.02
N Agents	285		103			

Table notes: The table displays the difference in means (Column (5)) between untreated (Columns (1)-(2)) and eventually treated (Columns (2)-(3)) agents in the last observed month preceding treatment introduction. Column (6) displays the corresponding p-value. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ denotes the work speed outcome, which is defined as the average handling time.

Table 1 shows no significant differences for age and gender, which yields no indication to control for those factors. Even though the same could be stated for tenure and being a temporary help agent, insignificant differences (on the 5% level) can be explained by the fact that, shortly before the introduction of the bonus, agents may have been hired on a temporary basis or did not (yet) meet the tenure criterion, and thus only became subject to treatment later in the treatment period. Still, both factors have a significant impact on treatment timing and should be considered as control variables. Finally, the table shows that agents who were eventually treated have more contractual hours and a greater number of evaluated calls. We will further explore this in Section 6.2.

Looking at service quality ($y_{i,t}^{QUAL}$) and work speed ($y_{i,t}^{QUAN}$) before the bonus was introduced, there was no difference in service quality between treated and untreated agents, but calls were 23 seconds shorter for treated agents, which translates into a *better* performance among agents who are eventually treated. Figures A2 and A3 in the Appendix further illustrate the development of service quality and quantity over time. We distinguish between high-skilled and low-skilled employees. For high-skilled agents, those that are subject to the bonus introduction show lower levels of service quality in March 2011. For low-skilled workers, the opposite is observable, which is consistent with observing no overall difference for $y_{i,t}^{QUAL}$ (Table 1). In terms of work speed, the graphs reveal that eventually treated agents report slightly lower levels of $y_{i,t}^{QUAN}$ (better performance) in March 2011, irrespective of the skill level.

4.3 Estimation Strategy

A well-established method to estimate the treatment effects of a policy intervention over time is the use of dynamic two-way fixed effects (TWFE) estimators of the following form:

$$y_{it}^{j} = \delta_i + \lambda_t + \alpha_1 X_{it}' + \sum_{l=-K}^{-2} \beta_l D_{it}^{l} + \sum_{l=0}^{L} \beta_l D_{it}^{l} + \epsilon_{it} \tag{3}$$

With y_{it}^{j} as the outcome of interest. j indicates whether the outcome is the directly incentivized quality outcome ($QUAL$) or the non-incentivized work speed outcome ($QUAN$). δ_i accounts for individual fixed effects, and λ_t capturing time-fixed effects. In our setting D_{it}^{l} represents the relative time period l of agent i with respect to the first time the bonus was applied ($l = 0$). Furthermore, we consider variables that may be related to both eligibility for the new bonus and service outcomes. Hence, X_{it}' comprises a vector of control variables including the number of planned working hours, tenure, sub-department dummies, the number of evaluated calls (per month), and a dummy for being a temporary help agent. β_l, with $l > 0$, then indicate the coefficients of interest and the effect of the bonus introduction on performance, conditioned on covariates, common time-trends, and individual time-invariant factors.

Because not all agents are directly eligible for the new bonus in April 2011, e.g. because they had short tenure, the treatment definition implies a staggered introduction of the bonus. In these cases, it has been shown that dynamic two-way fixed effects (TWFE) may produce biased results under staggered adoption (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Roth et al., 2022). Besides comparisons between treated and not-yet-treated units, TWFE estimates also make comparisons between units that have both been treated. The latter comparison, however, can lead to negative weighting and even lead to estimated treatment effects having the opposite sign in extreme cases (Roth et al., 2022). Hence, to overcome this limitation, we additionally use the method proposed by Sun and Abraham (2021), allowing for staggered treatment and treatment effect heterogeneity. The estimation strategy follows a similar rationale as dynamic TWFE estimators (Equation (3)), but uses only agents that are never subject to the bonus as a control group, in order to allow valid comparisons. With that, we obtain sensible treatment effects constructed as the weighted average of coefficients for each cohort and each relative time after or before the treatment.[11]

Counterfactual In our setting, not only did the treatment group receive a new (individual) bonus, but the control group also began to benefit from a new (group-related) bonus scheme in April 2011. As a result, the average treatment effect represents the effect of an individual bonus that exceeds the effect of a bonus based on group performance and our findings may underestimate the true impact of instituting an individual bonus scheme.

Parallel trends Dynamic estimates based on Sun and Abraham (2021) rely on parallel trends and no anticipation assumptions. Besides visually inspecting event-study graphs in that set-up, we address those assumptions using other estimators based on Callaway and Sant'Anna (2021) and Gardner (2022) that are sensible under staggered treatment and treatment effect heterogeneity in Section 6.

5 Results

5.1 Main Results

Table 2 displays the main results with respect to service quality $y_{i,t}^{QUAL}$ (Columns (1)-(4)) and work speed $y_{i,t}^{QUAN}$ (Columns (5)-(8)). Furthermore, we report TWFE estimates (Columns (1), (2), (5) and (6)) and estimates based on Sun and Abraham (2021) (Columns (3), (4), (7) and (8)). We present one specification excluding (uneven columns) and one including (even columns) control variables to see if the treatment effects are ro-

[11] For a detailed overview of the method and its assumptions we refer to Sun and Abraham (2021) and Roth et al. (2022).

bust. The results from the entire sample show no significant treatment effects in any of the specifications for $y_{i,t}^{QUAL}$. The implementation of the bonus payment had no effect on overall service quality. Furthermore, the table suggests that the control variables have no effect on service quality either.

Table 2: The Effect of the Bonus Introduction on Service Quality and Work Speed

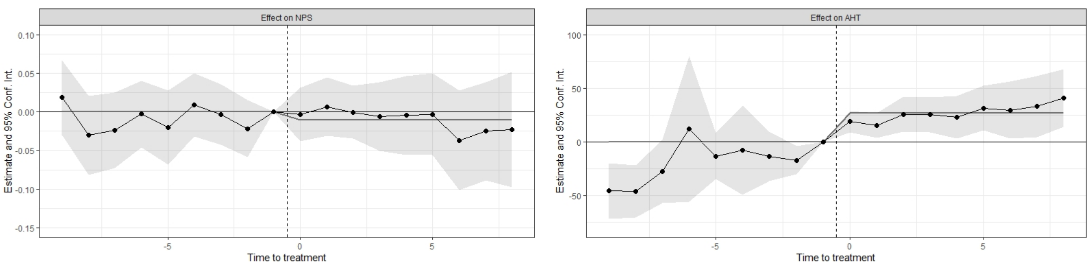
Dependent Variable:	Service quality (y_{it}^{QUAL})				Work speed (y_{it}^{QUAN})			
Model:	TWFE		Sun and Abraham (2021)		TWFE		Sun and Abraham (2021)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Performance Pay	0.002	-0.001	0.000	-0.008	30.6***	36.4***	18.5***	25.5***
	(0.011)	(0.022)	(0.017)	(0.018)	(8.06)	(8.03)	(6.91)	(7.70)
Hours planned		0.0005		0.0003		0.359		0.473
		(0.002)		(0.002)		(0.463)		(0.466)
Tenure		0.016		0.019		-4.54		-2.74
		(0.020)		(0.020)		(6.22)		(6.05)
Evaluated calls		0.000		0.000		-0.728***		-0.799***
		(0.0004)		(0.0004)		(0.152)		(0.165)
Temporary Agent		-0.009		-0.003		22.4***		18.6***
		(0.021)		(0.024)		(7.44)		(7.09)
Department dummies	No	Yes	No	Yes	No	Yes	No	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,517	3,517	3,517	3,517	3,517	3,517	3,517	3,517
R^2	0.261	0.280	0.529	0.299	0.529	0.567	0.564	0.598
Mean Dependent Variable	0.398				342.177			
N Agents	388				388			

Table notes: The results display average treatment effects of the bonus introduction on $y_{i,t}^{QUAL}$ and $y_{i,t}^{QUAN}$ based on standard TWFE (Columns (1) to (3)) and Sun and Abraham (2021) dynamic estimators (Columns (3) and (4). The model sub-sequentially adds control variables and individual FE. The sample includes the months between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

For the non-incentivized work speed $y_{i,t}^{QUAN}$, the results display a positive treatment effect. The bonus introduction led to an increase in $y_{i,t}^{QUAN}$, i.e., longer calls, which translates into weaker agent performance. This result could be explained by a shift in effort from non-incentivized work speed to incentivized service quality. Furthermore, the treatment effects change when accounting for control variables. In both the TWFE and estimates based on Sun and Abraham (2021), the inclusion of additional control variables yields higher treatment coefficients (and hence lower performance). One reason for that result may be the fact that agents become eligible for treatment only if they are not employed on a temporary basis and temporary help agents perform worse in terms of work speed. Hence, if we do not control for being a temporary help agent, treatment effects also capture the effect of workers not being temporarily employed. Therefore, the initial estimates may *underestimate* the overall treatment effect in terms of work speed. Besides the importance of control variables, the coefficients are different depending on the model specification, as estimates based on Sun and Abraham (2021) are lower than standard TWFE estimates. The staggered adoption and treatment effect heterogeneity may lead to an overestimation of coefficients in dynamic TWFE set-ups. Therefore, for the rest of the analysis, we use the estimator based on Sun and Abraham (2021) as the main specification and - given the sensitivity of the coefficients - account for differences in control variables (Columns (4) and (8) of Table 2).

Parallel trends are a crucial assumption of the estimator based on Sun and Abraham (2021). Hence, we proceed by visually inspecting the pre-trend period of the main estimations. Figure 2 shows the respective event studies. For the quality outcome y_{it}^{QUAL} (Sub-figure (a)), the results show no significant pre-trends in any of the relative pre-periods. For the work speed measure y_{it}^{QUAN} (Sub-figure (b)), some periods seem to violate the parallel trends assumptions. However, in the seven months leading up to the treatment, the coefficients do not differ statistically from zero. Still, to see whether this result is robust, we use alternative event-study designs and model specifications in subsection 6.

Figure 2: Event Study of the Effect of the Bonus Introduction (Full Sample)



(a) Service quality (y_{it}^{QUAL}) (b) Work speed (y_{it}^{QUAN})

Figure notes: Event Study Analysis based on Sun and Abraham (2021) showing treatment effects on $y_{i,t}^{QUAL}$ (a) and $y_{i,t}^{QUAN}$ (b) for each relative treatment. The sample includes the time between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the service quality outcome represented by the agent's average handling time.

5.2 Skill Heterogeneity

Overall, the results so far imply no treatment effects on the incentivized service quality outcome y_{it}^{QUAL} and negative effects on work speed y_{it}^{QUAN}, i.e. longer calls. However, if workers differ in their costs of effort to reach performance thresholds, the results may vary depending on the level of skill. In Table 3, we re-estimate the main specification but distinguish between agents that are relatively low-skilled (Columns (1) and (3)) versus high-skilled (Columns (2) and (4)). The results show an important heterogeneity: We observe a positive treatment effect for those agents that are relatively high-skilled. As y_{it}^{QUAL} is defined from 0 to 1, being subject to the bonus payment is associated with an increase in quality of approximately 5.7 percentage points. With respect to the pre-treatment mean, this can be translated into an increase in the performance of high-skilled workers by approximately 11% (0.057/0.501). On the other hand, relatively low-skilled agents display a negative average treatment effect on the 10% level, which translates into a decrease in performance by 14% (0.040/0.295).

For the non-incentivized outcome y_{it}^{QUAN}, which is based on average handling time, Table 3 reveals that the overall effects on work speed are driven by agents that are relatively low-skilled. Given the pre-introduction mean of approximately 333 seconds for low-skilled workers, the bonus introduction led to a decrease in y_{it}^{QUAN} by 9% (30.5/333). The coefficient for high-skilled agents remains statistically insignificant.

Again, to address the parallel trends assumption, we proceed by visually inspecting the pre-trend period. For y_{it}^{QUAL} Figure 3 displays no significant pre-trend irrespective of the level of skill. For the work speed measure y_{it}^{QUAN}, similar to the full sample, Figure 4 shows some periods further away from treatment violate the parallel trends assumptions for both high-skilled and low-skilled employees, which will be further addressed in Section 6.

Besides the visual inspection of potential pre-trends, event-study graphs allow the inspection of dynamic treatment effects. In terms of y_{it}^{QUAL} one can observe that the treatment effect for high-skilled agents shows no apparent pattern. On the other hand, low-skilled agents' effects on y_{it}^{QUAN} slightly increase, with a small initial response in the two months after the treatment. Furthermore, especially for the low-skilled agents, the

Table 3: The Effect of the Bonus Introduction - Skill Differences

Dependent Variables:	Service Quality (y_{it}^{QUAL})		Work Speed (y_{it}^{QUAN})	
Sub-Sample	Low-skilled	High-skilled	Low-skilled	High-skilled
	(1)	(2)	(3)	(4)
Performance Pay	-0.040*	0.057**	30.5***	16.5
	(0.023)	(0.027)	(9.81)	(11.1)
Control variables	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	1,777	1,740	1,777	1,740
R^2	0.370	0.288	0.601	0.694
Mean dependent Variable	0.295	0.501	333.452	350.903
N Agents	194	194	194	194

Table notes: The results display average treatment effects of the bonus introduction on $y_{i,t}^{QUAL}$ and $y_{i,t}^{QUAN}$ depending on the level of skill. Columns (1) and (3) show treatment effects for low-skilled agents, while Columns (2) and (4) exhibit effects for high-skilled agents. The sample includes the months between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. Standard errors are clustered at the agent level. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

treatment effects are sensitive to the choice of the reference period (t = *-1*), as we observe a sharp increase for both y_{it}^{QUAL} and y_{it}^{QUAN} in that period. The sensitivity of the results to the choice of reference periods will be further discussed in Subsection 6.

In Table A1, we split our sample based on the tertile of the ability scores (rather than the median) to further highlight effect heterogeneity at the tails of the distribution. In that specification, we refer to agents as 'low', 'medium', and 'high' skilled. The findings - even though the sample restriction reduces statistical power - show that for service quality as an outcome, negative treatment effects are found only for those at the lower end of the distribution, while positive effects are found only for high-skilled agents. For the latter, the positive effect is even stronger compared to the median split (0.072 versus 0.057). Furthermore, the negative effects on service quality at the bottom of the ability distribution increase in both economic and statistical significance. Medium-skilled employees show no significant treatment effects. Negative effects on work speed, are driven by low and medium-skilled employees, while high-skilled employees display no response to the bonus introduction for that outcome.

Figure 3: The Effect of the Bonus Introduction on Service Quality (y_{it}^{QUAL}) - Event Study

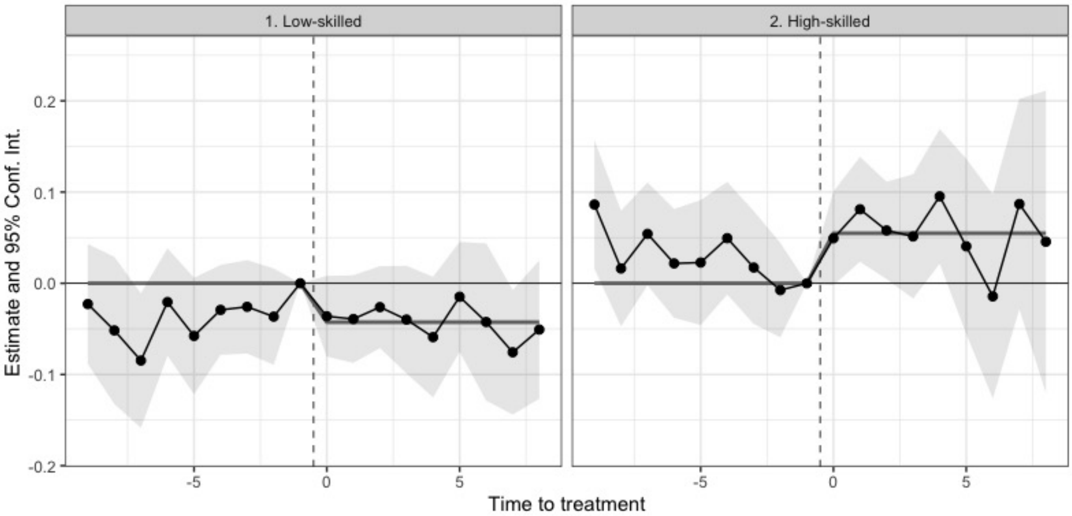


Figure notes: Event Study Analysis based on Sun and Abraham (2021) showing treatment effects on $y_{i,t}^{QUAL}$ for each relative treatment period. The sample includes the time between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS).

As explained in Section 3.1, the NPS (or y_{it}^{QUAL}) consists of three service quality components: Bad (0-6), medium (7-8) and good (9-10). Hence, we can further disentangle the effect of the bonus introduction on the three service quality outcomes, to see where the main result stems from. To do so, in Table A2 we use the share of bad (Columns (1)-(3)), medium (Columns (4)-(6)) and good (Columns (7)-(9)) calls as dependent variable. For the full sample and low-skilled employees, we find no significant effects on all three outcomes, while the initial effect for high-skilled agents on y_{it}^{QUAL} seems to be driven mainly by a higher share of good calls with an increase of around 22% compared to the pre-treatment mean.

5.3 Performance Bonuses and Overtime

While for quantitative performance targets increasing working hours may be one channel that explains higher performance, it can be even harmful when agents want to achieve higher quality, as extended working hours can lead to reduced concentration and, consequently, a decline in the quality of calls. Additionally, especially for high-skilled workers,

Figure 4: The Effect of the Bonus Introduction on Work Speed (y_{it}^{QUAN}) - Event Study

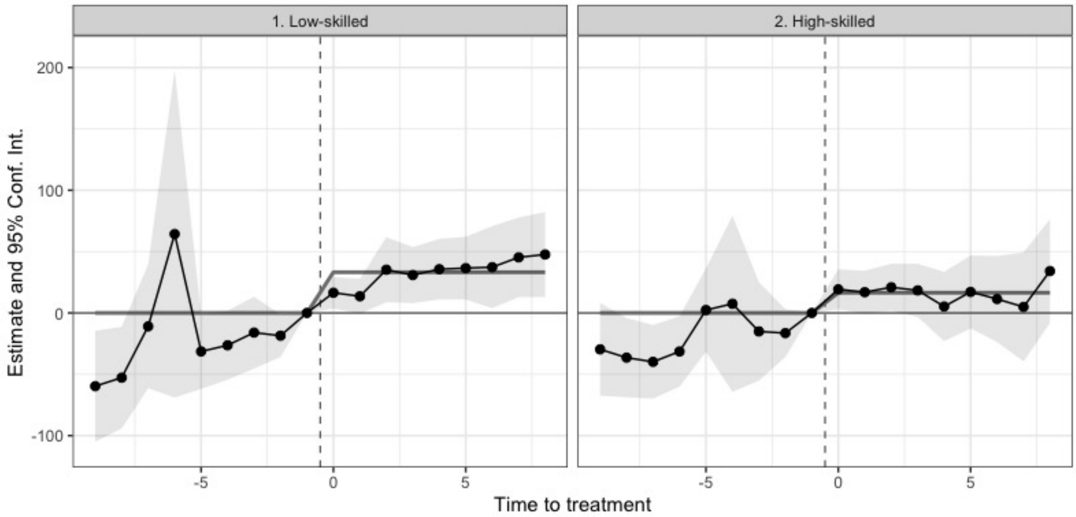


Figure notes: Event Study Analysis based on Sun and Abraham (2021) showing treatment effects on $y_{i,t}^{QUAN}$ for each relative treatment period. The sample includes the time between October 2010 and December 2011. $y_{i,t}^{QUAN}$ is the work speed outcome expressed in the agent's average handling time.

performance bonuses can function as a substitute for paid overtime, yielding negative effects on hours of work.[12]

Information on actual working hours is provided for agents in or after December 2010, which reduces the number of observations to 2,396. We focus on overtime hours, as they may better capture the (immediate) agents' behavioral response to the bonus, while adjustments of contractual working hours may take longer transition periods. We construct overtime as the difference between contractual work hours and the actual hours worked. On average, both high-skilled and low-skilled agents work approximately 30 hours per week under contract in our sample. In the last observed month before the bonus introduction, high-skilled (low-skilled) workers performed on average 5.06 (4.61) hours fewer than their contractual working hours. Regardless of other control variables, this gap shrank to 4.42 (3.64) hours per week in the nine months after the introduction.

Table 4 displays the estimated coefficients when regressing the average hours of overtime (per week) on the same set of control variables as in Table 2, as they may correlate with overtime and the probability of being subject to treatment. The results

[12]Recent survey evidence on performance-related pay and working hours (Artz and Heywood, 2023; Green and Heywood, 2023) show an increase in working hours in response to such incentives. However, the nature of performance pay and whether it incentivizes quality or quantity remains unclear in those studies given the questionnaire design.

show, that on average, workers reduced weekly overtime by 1.51 hours in response to the bonus payment, with low-skilled agents showing a larger and statistically significant decline.

Table 4: The Effect of the Bonus Introduction on Overtime

Dependent Variable:	Overtime Hours (weekly)		
Sample	Full sample	Low-skilled	High-skilled
	(1)	(2)	(3)
Performance Pay	-1.51**	-1.53*	-1.11
	(0.611)	(0.901)	(0.890)
Hours planned	-0.218***	-0.177***	-0.258***
	(0.036)	(0.043)	(0.060)
Tenure	0.854	1.07	-0.954
	(0.530)	(0.716)	(1.21)
Evaluated calls	0.039***	0.028**	0.053***
	(0.008)	(0.011)	(0.012)
Temporary Agent	0.202	0.031	0.789
	(0.427)	(0.566)	(0.659)
Department dummies	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	2,396	1,204	1,192
R^2	0.546	0.542	0.596
Mean dependent Variable	-4.84	-4.61	-5.06
N Agents	303	152	151

Table notes: The results display average treatment effects of the bonus introduction on weekly overtime hours. The sample includes the months between October 2010 and December 2011. Overtime hours are defined as the difference between actual and contractual working hours. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A4 shows that the parallel trends assumption consistently holds for all the sub-samples and all the periods. Overall, the findings suggest that, in contrast to previous findings on performance pay and working hours, incentivizing quality may cause lower working hours. This result can be interpreted such that workers have a tendency to avoid long hours in order to potentially avoid negative evaluations by customers, which have been documented for other performance outcomes (Collewet and Sauermann, 2017).

6 Robustness

6.1 Model Specification

Anticipation of the introduction One might be concerned that workers anticipate and have a strategic incentive to underperform *before* the introduction of the new bonus system to have lower performance thresholds that are easier to reach (the Ratchet effect; Weitzman, 1980; Charness et al., 2011). From an individual agent's perspective, it would be a risky strategy to consistently underperform for an unspecified in an environment where performance is easy to observe along multiple dimensions and at high frequency. That agents collectively underperform to lower performance thresholds seems difficult in terms of coordination. Since we are not interested in the level of the performance threshold *per se*, lower performance in periods just before the introduction of the bonus pay should be visible in the event-study version of our results. However, especially in the incentivized outcome y_{it}^{QUAL}, we do not observe a systematic decline before the treatment, as the parallel trends assumption is not violated.[13] If ratchet effects should nonetheless be at work in our setting, this should instead lead to an underestimation given that we use a long period before the introduction.

Still, to emphasize this point and the sensitivity of our results to the reference period used, we use both *t-1* and *t-2* as baseline categories in our event study estimates (see Table A3, Figure A5 and Figure A6). With that, our baseline results amplify: The effect of the bonus introduction on y_{it}^{QUAL} increases from 0.057 to 0.062 for the high-skilled and the effect on y_{it}^{QUAN} from 30.5 to 40.9 for the low-skilled agents.[14] Again, this implies that our initial estimates are rather conservative estimates compared to the true treatment effect. Furthermore, when looking at the event-study designs, there is no indication that the parallel trends assumption is violated, particularly in the 6 months preceding the treatment. In the following section, we will investigate potential violations of parallel trends further.

[13] If any, we observe a significant increase in performance for low-skilled agents, just before the bonus was introduced.

[14] Note that the effects on y_{it}^{QUAN} for the high-skilled agents also turn statistically significant with the adjusted reference period.

Violation of parallel trends Apart from visually inspecting potential violations of parallel trends in the framework of Sun and Abraham (2021), we use two alternative specifications that allow for staggered treatment and treatment effect heterogeneity. First, Gardner (2022)'s alternative approach provides a two-stage Difference-in-Difference estimator.[15] Second, using Callaway and Sant'Anna (2021)'s estimator allows the inclusion of inverse probability weighting, instead of relying on a regression-based design to calculate the average treatment effects, depending on co-variates.[16]. Similar to dynamic TWFE approaches like Sun and Abraham (2021), both estimators provide sensible results assuming parallel trends and no anticipation.

Significantly, the results displayed in Figures A7 and A8 show no violation of the parallel trends assumption in the six months leading up to treatment in any of the specifications. Again, violations that happen earlier, are unlikely to be related to the introduction of the treatment. Furthermore, the findings displayed in Table A4 confirm a positive treatment effect in terms of y_{it}^{QUAL} for the high-skilled and a negative treatment effect on y_{it}^{QUAN} for the low-skilled.[17]

Peer effects The bonus introduction may influence untreated agents for a variety of reasons: To start, the characteristics of teams may differ in terms of competitiveness, which may cause peer effects on performance. In competitive teams, if treated agents begin to improve their performance in response to the bonus, non-treated agents may adjust their effort to avoid falling behind. While we cannot directly account for intra-team competition, we can account for time-invariant differences in team composition by incorporating team-fixed effects into our main specification. The results displayed in Table A5 (Panel A) show that the inclusion of team-fixed effects does not affect our results.[18]

[15]The estimator first identifies group and period effects from the untreated sample observations and then identifies average treatment effects by comparing treated and untreated outcomes after removing such effects.

[16]We exclude the sub-Department dummies in that specification, as its inclusion leads to co-linearity issues in the framework of inverse probability weighting.

[17]The effect on y_{it}^{QUAN} for low-skilled agents in the framework of Callaway and Sant'Anna (2021) is only significant at the 11%-level. Given the small (sub-)sample size, this is most likely a problem with the assignment of feasible weights, as also the significance in other samples decreases in comparison to estimators based on Sun and Abraham (2021) and Gardner (2022).

[18]The number of observations slightly decreases, as the team id is not available for all observations.

Additionally, the share of treated individuals per team may put peer pressure on untreated agents, as for instance, the communication about the newly introduced bonus is higher in teams that are more severely affected, which may affect the effort of untreated agents. To shed light on that issue, we include the average share of treated agents in the post period ($\overline{T_{post}}$) as a control variable in Table A5 (Panel B). The treatment effects are qualitatively the same as in the baseline specification, and the results for y_{it}^{QUAN} indicate that teams with a large share of agents affected by the bonus work faster on average.

6.2 Variable Choice and Sample Trimming

Number of evaluated calls A potential issue with the NPS score as an outcome variable is its variation in the number of calls evaluated. For example, if the number of evaluated calls increases systematically for high-skilled agents versus low-skilled agents after the bonus introduction, this may result in a significant difference in the variability of y_{it}^{QUAL}. This worry is mitigated by the fact that management is not able to affect *which* customers are evaluated, as callbacks occurred randomly. We consider the number of evaluated calls as the dependent variable, and the results in Table A6 show that there is no statistically significant difference between high-skilled and low-skilled agents in terms of evaluated calls.

Alternative measure of skill The measure of skill used in this study is based only on agents observed between October 2010 and March 2011. However, excluded agents that are not observed in that period, may have different properties. To recover those agents, we define a broader skill measure that relies on the observed service quality in all the quarters before the introduction.[19] Compared to the main specification, the sample size increases by around 180 observations (or 5%). The results (Table A7) confirm a positive treatment effect for high-skilled agents in terms of y_{it}^{QUAL} and negative treatment effects on y_{it}^{QUAN} for low-skilled employees. However, given the adjusted skill measure, we do not find significant negative treatment effects on the low-skilled in terms of service quality.

[19]The procedure follows the same procedure as in Section 3.2, but uses the whole pre-period for estimation.

Additionally, it seems that also high-skilled agents performed significantly worse in work speed after the bonus was introduced.[20]

Sample trimming The sample contains a large number of agents (139) who were only observed before the bonus was introduced. Because these agents affect the difference between the treatment and control groups in the pre-period, they also affect the overall average treatment effect. If relatively 'bad' agents are in the control group before the bonus was introduced but not later, then the pre-treatment difference may be overestimated and post-treatment differences would be even larger if those agents had not left the firm. This could imply that the overall effect is biased downward. On the other hand, if 'good' agents leave the firm, the opposite is true.[21] To account for the issue, we re-estimate our baseline result, conditional on agents being observed in months before *and* after the introduction of the bonus. The results in Table A8 indicate similar coefficients as in our baseline estimates and the data's unbalanced structure does not drive our main findings.

7 Conclusion

This study analyzes the effect of the introduction of individual performance pay on worker performance using unique data on agents working in the call center of a multi-national telephone company. The data contain qualitative as well as quantitative performance information before and after the introduction of a performance bonus. While the bonus pay was based merely on service quality, one of the performance outcomes, work speed, was not incentivized.

To analyze dynamic treatment effects and the effectiveness of the bonus introduction, we employ a variety of estimation strategies and model specifications. Our main results consistently show no evidence that agents react to incentives set by management and increase their performance in terms of service quality to get a monetary bonus. This

[20]We perform additional robustness tests, defining skill based on only the quarter leading up to the treatment and find results that confirm those findings. The results can be obtained from the authors upon request.

[21]We find no evidence in the data that relatively good or bad agents leave the firm, as around 50% of them were high-skilled (low-skilled). However, differences within the sub-samples may occur.

differs from the majority of previous studies that examine the effectiveness of individual performance pay and worker performance at the firm level, which may be due to the qualitative nature of the incentivized outcome. The non-incentivized (quantitative) performance outcome, on the other hand, decreased in response to the introduction of the bonus. However, effect heterogeneity is important in both incentivized and non-incentivized outcomes: High-skilled agents increased service quality by approximately 11% in response to the newly introduced bonus payment and show, if any, only a small negative effect on work speed. Low-skilled agents, on the other hand, did not improve service quality and even decreased work speed by 9%. Finally, contrary to recent empirical studies, the current study finds that under qualitative performance incentives, agents reduce overtime hours in response to individidual performance pay.

A conclusion that can be drawn from the findings is that for low-skilled workers, the introduction of individual performance pay may not lead to an increase in performance if the targets are hard to achieve. In such cases, the incentive to improve productivity may not be strong enough, leading to little or no performance gains. For these agents, focusing on performance gains that are hard to achieve might even backfire and create negative spillover effects on their non-incentivized performance outcomes. As a result of these negative consequences, managers should carefully consider the viability of individual bonus schemes before implementing them. Furthermore, performance bonuses can even dampen workers' willingness to work overtime as excess working hours can harm the quality of calls.

References

Al-Ubaydli, O., S. Andersen, U. Gneezy, and J. A. List (2015): "For Love or Money? Comparing the Effects of Non-pecuniary and Pecuniary Incentive Schemes in the Workplace," *Southern Economic Review*, 81, 538–561.

Angrist, J. and V. Lavy (2009): "The Effects of High Stakes High School Achievement Awards: Evidence from a randomized Trial," *American Economic Review*, 99, 1384–1414.

Artz, B. and J. S. Heywood (2023): "Performance pay and work hours: US survey evidence," *Oxford Economic Papers*.

Asch, B. J. (1990): "Do Incentives Matter? The Case of Navy Recruiters," *Industrial and Labor Relations Review*, 43, pp. 89S–106S.

Azmat, G. and N. Iriberri (2010): "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students," *Journal of Public Economics*, 94, 435–452.

Bandiera, O., I. Barankay, and I. Rasul (2007): "Incentives for Managers and Inequality Among Workers: Evidence From a Firm-Level Experiment," *Quarterly Journal of Economics*, 122, 729–773.

Barrero, J. M., N. Bloom, and S. J. Davis (2023): "The Evolution of Working from Home," unpublished manuscript, Stanford University.

Battiston, D., J. Blanes i Vidal, and T. Kirchmaier (2021): "Face-to-Face Communication in Organisations," *Review of Economic Studies*, 88, 574–609.

Bradler, C., S. Neckermann, and A. J. Warnke (2019): "Incentivizing Creativity: A Large-scale Experiment with Performance Bonuses and Gifts," *Journal of Labor Economics*, 37, 793–851.

Bryan, M. and A. Bryson (2016): "Has Performance Pay Increased Wage Inequality in Britain?" *Labour Economics*, 41, 149–161.

Cadsby, C. B., F. Song, and F. Tapon (2007): "Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation," *Academy of Management Journal*, 50, 387–405.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-differences with Multiple Time Periods," *Journal of Econometrics*, 225, 200–230.

Camargo, B., F. Lange, and E. Pastorino (2022): "On the Role of Learning, Human Capital, and Performance Incentives for Wages," NBER Working Paper 30191, National Bureau of Economic Research.

Charness, G., P. Kuhn, and M. C. Villeval (2011): "Competition and the Ratchet Effect," *Journal of Labor Economics*, 29, 513–547.

Collewet, M. and J. Sauermann (2017): "Working hours and productivity," *Labour Economics*, 47, 96–106.

De Grip, A. and J. Sauermann (2012): "The Effects of Training on Own and Co-Worker Productivity: Evidence from a Field Experiment," *Economic Journal*, 122, 376–399.

De Grip, A., J. Sauermann, and I. Sieben (2016): "The Role of Peers in Estimating Tenure-Performance Profiles: Evidence from Personnel Data," *Journal of Economic Behavior & Organization*, 126, 39–54.

DeVaro, J. (2022): "Performance Pay, Working Hours, and Health-related Absenteeism," *Industrial Relations*, 61, 327–352.

Dohmen, T. and A. Falk (2011): "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender," *American Economic Review*, 101, 556–90.

Eriksson, T. and M. C. Villeval (2008): "Performance-pay, Sorting and Social Motivation," *Journal of Economic Behavior & Organization*, 68, 412–421.

Eurofund (2020): "Employee Monitoring and Surveillance: The Challenges of Digitalisation," *Publications Office of the European Union, Luxembourg.*

Franceschelli, I., S. Galiani, and E. Gulmez (2010): “Performance Pay and Productivity of Low- and High-ability Workers,” *Labour Economics*, 17, 317–322.

Gardner, J. (2022): “Two-stage Differences in Differences,” *arXiv preprint arXiv:2207.05943*.

Gielen, A. C., M. J. M. Kerkhofs, and J. C. van Ours (2010): “How Performance related Pay Affects Productivity and Employment,” *Journal of Population Economics*, 23, 291–301.

Gneezy, U. and A. Rustichini (2000): “Pay Enough or Don’t Pay at All,” *Quarterly Journal of Economics*, 115, 791–810.

Gompers, P., A. Kovner, J. Lerner, and D. Scharfstein (2010): “Performance persistence in entrepreneurship,” *Journal of Financial Economics*, 96, 18–32.

Green, C. P. and J. S. Heywood (2023): “Performance Pay, Work Hours and Employee Health in the UK,” *Labour Economics*, 84, 102387.

Grund, C. and D. Sliwka (2010): “Evidence on Performance Pay and Risk Aversion,” *Economics Letters*, 106, 8–11.

Havranek, T., P. Cala, Z. Irsova, J. Matousek, and J. Novak (2022): “Financial Incentives and Performance: A Meta-Analysis of Economics Evidence,” CEPR Discussion Papers 17680.

Heywood, J. S., X. Wei, and G. Ye (2011): “Piece Rates for Professors,” *Economics Letters*, 113, 285–287.

Holmstrom, B. and P. Milgrom (1991): “Multitask Principal-agent Analyses: Incentive Contracts, Asset Ownership and Job Design,” *Journal of Law, Economics and Organization*, 7, 24–52.

Hong, F., T. Hossain, J. A. List, and M. Tanaka (2018): “Testing the Theory of Multitasking: Evidence from a natural Field Experiment in Chinese Factories,” *International Economic Review*, 59, 511–536.

Kato, T. and P. Shu (2008): "Performance Spillovers and Social Network in the Workplace: Evidence from Rural and Urban Weavers in a Chinese Textile Firm," IZA Discussion Paper 3340, Institute of Labor Economics.

Keiningham, T. L., B. Cooil, T. W. Andreassen, and L. Aksoy (2007): "A Longitudinal Examination of Net Promoter and Firm Revenue Growth," *Journal of Marketing*, 71, 39–51.

Kleine, M. (2021): "No Eureka! Incentives Hurt Creative Breakthrough Irrespective of the Incentives' Frame," Max Planck Institute for Innovation & Competition Research Paper 21-15.

Kowalski, A. M. (2019): "Is There Folly When Worker A is More Productive Than Worker B?: Examining Heterogeneous Responses to Individual and Group Performance Pay," Ph.D. thesis, Massachusetts Institute of Technology.

Lavy, V. (2009): "Performance Pay and Teachers' Effort, Productivity, and Grading Ethics," *American Economic Review*, 99, 1979–2011.

Lazear, E. P. (2000): "Performance Pay and Productivity," *American Economic Review*, 90, 1346–1361.

Lemieux, T., W. B. Macleod, and D. Parent (2009): "Performance Pay and Wage Inequality," *Quarterly Journal of Economics*, 124, 1–49.

Liu, X. and R. Batt (2007): "The Economic Pay-Offs to Informal Training: Evidence from Routine Service Work," *Industrial and Labor Relations Review*, 61, 75–89.

Maestas, N., K. J. Mullen, D. Powell, T. Von Wachter, and J. B. Wenger (2017): "Working Conditions in the United States: Results of the 2015 American Working Conditions Survey," Tech. rep., RAND Corporation, Santa Monica, CA.

Manthei, K., D. Sliwka, and T. Vogelsang (2021): "Performance Pay and Prior Learning—Evidence from a Retail Chain," *Management Science*, 67, 6998–7022.

——— (2022): "Information, Incentives, and Attention: A Field Experiment on the Interaction of Management Controls," *The Accounting Review*, forthcoming.

O'Halloran, P. L. (2012): "Performance Pay and Employee Turnover," *Journal of Economic Studies*, 39, 653–674.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2022): "What's Trending in Difference-in-differences? A Synthesis of the Recent Econometrics Literature," *arXiv preprint arXiv:2201.01194*.

Sauermann, J. (2023): "Worker Reciprocity and the Returns to Training: Evidence from a Field Experiment," *Journal of Economics & Management Strategy*, 32, 543–557.

Shearer, B. (2004): "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment," *Review of Economic Studies*, 71, 513–534.

Shi, L. (2010): "Incentive Effect of Piece-Rate Contracts: Evidence from Two Small Field Experiments," *B.E. Journal of Economic Analysis & Policy*, 10, 61.

Sun, L. and S. Abraham (2021): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225, 175–199.

Taylor, E. S. (2022): "Employee Evaluation and Skill Investments: Evidence from Public School Teachers," NBER Working Paper 30687, National Bureau of Economic Research.

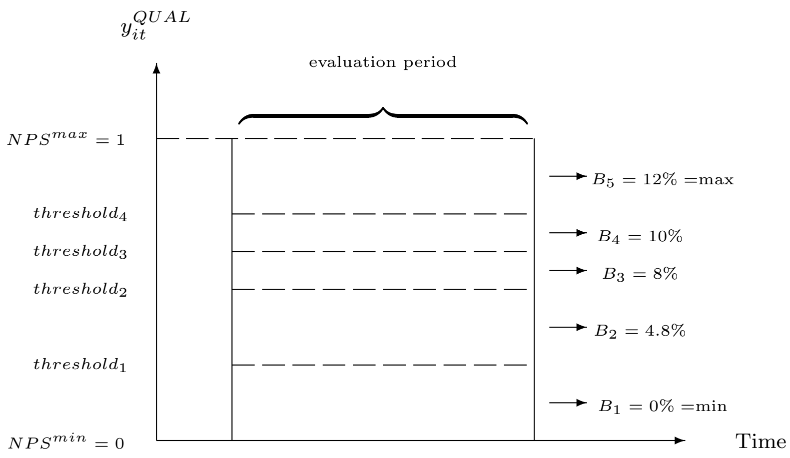
Unger, O., A. Szczesny, and M. Holderried (2020): "Does Performance Pay increase Productivity? Evidence from a Medical Typing Unit," *Management Accounting Research*, 47, 100649.

Weidmann, B. and D. J. Deming (2021): "Team Players: How Social Skills Improve Team Performance," *Econometrica*, 89, 2637–2657.

Weitzman, M. L. (1980): "The 'Ratchet Principle' and Performance Incentives," *The Bell Journal of Economics*, 302–308.

A Appendix

Figure A1: Design of the Bonus System



Note: B_j denotes bonus levels as a percentage of an agent's gross wage in the bonus quarter when achieving a performance level of $threshold_j$.

Figure A2: Evolution of Service Quality y_{it}^{QUAL} over Time

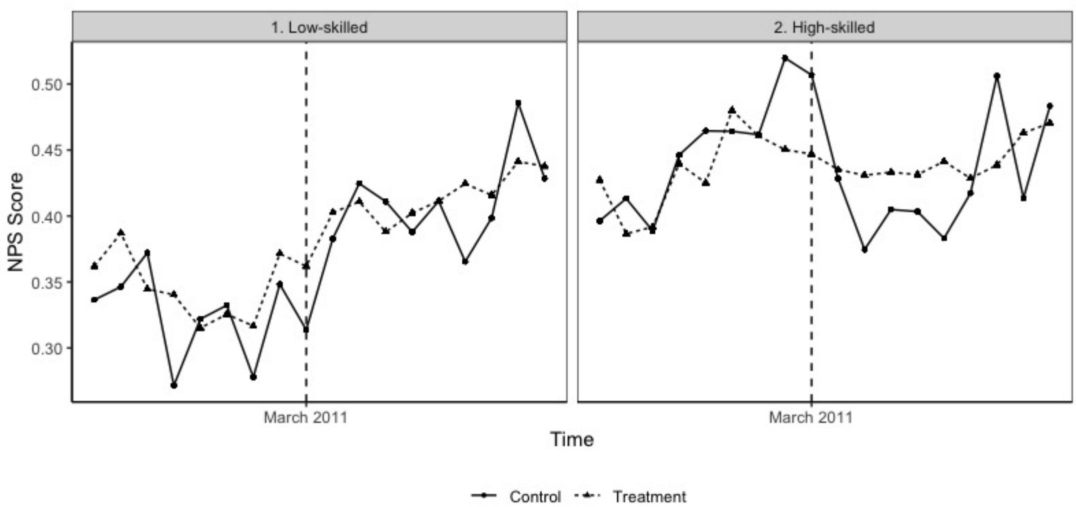


Figure notes: This figure displays $y_{i,t}^{QUAL}$ over time for treated and non-treated agents. The left-hand side shows the results for low-skilled and the right-hand sight for high-skilled agents. The last month before the treatment is indicated by the vertical dashed line. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS).

Figure A3: Evolution of Work Speed y_{it}^{QUAN} over Time

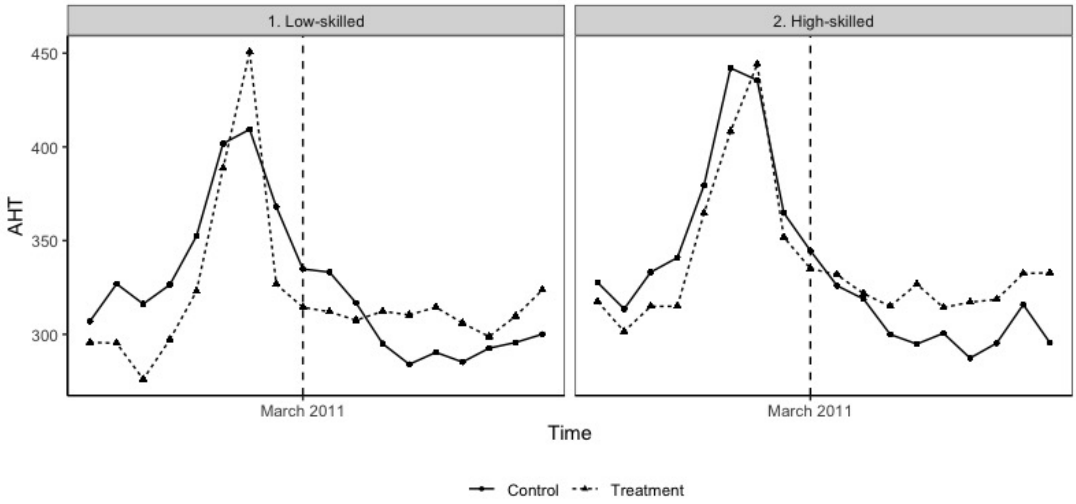


Figure notes: This figure displays y_{it}^{QUAN} for treated and non-treated agents over time. The left-hand side shows the results for low-skilled and the right-hand sight for high-skilled agents. The last month before the treatment is indicated by the vertical dashed line. $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time.

Figure A4: Event Studies of the Effect of the Bonus Introduction on Overtime

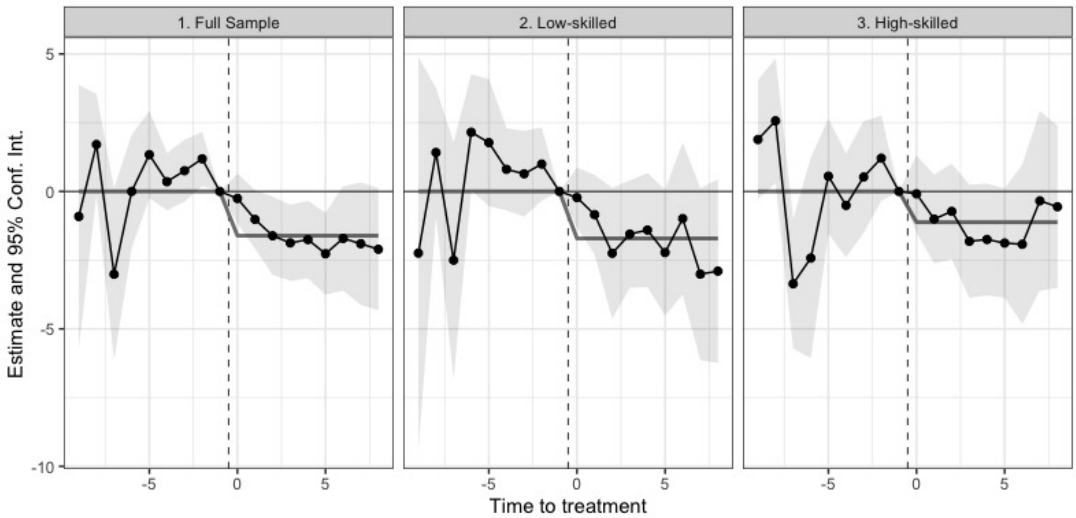


Figure notes: Event Study Analysis based on Sun and Abraham (2021) showing treatment effects on weekly overtime hours for each relative treatment period. The sample includes the time between October 2010 and December 2011. Overtime hours are defined as the difference between actual and contractual working hours.

Figure A5: The Effect of the Bonuses on Service Quality (y_{it}^{QUAL}) - Event Studies with Adjusted Reference Period

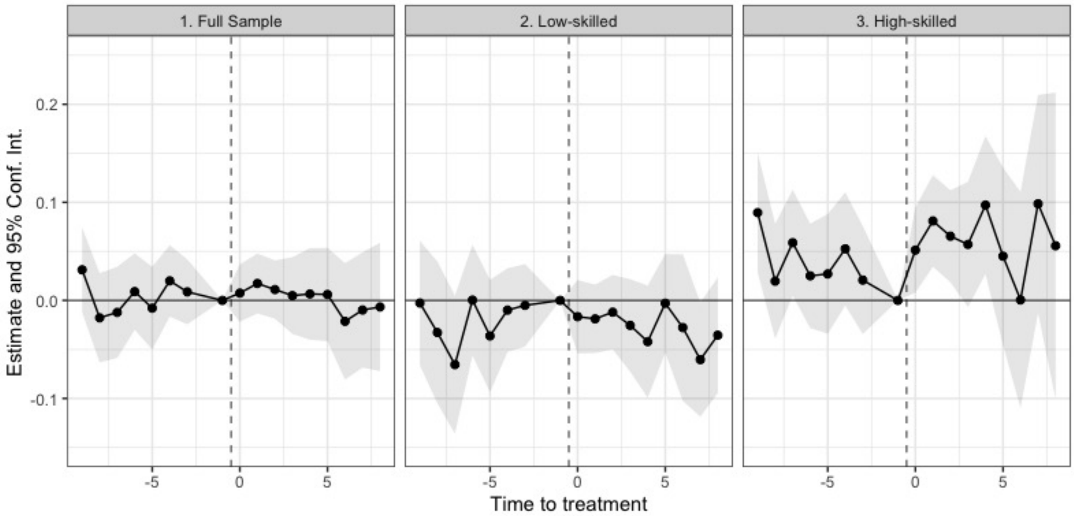


Figure notes: Event Study Analysis based on Sun and Abraham (2021) showing treatment effects on $y_{i,t}^{QUAL}$ for each relative treatment period. The sample includes the time between October 2010 and December 2011. The estimates use *t-1* and *t-2* as reference category. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS).

Figure A6: The Effect of the Bonuses on Work Speed (y_{it}^{QUAN}) - Event Studies with Adjusted Reference Period) - Event Study

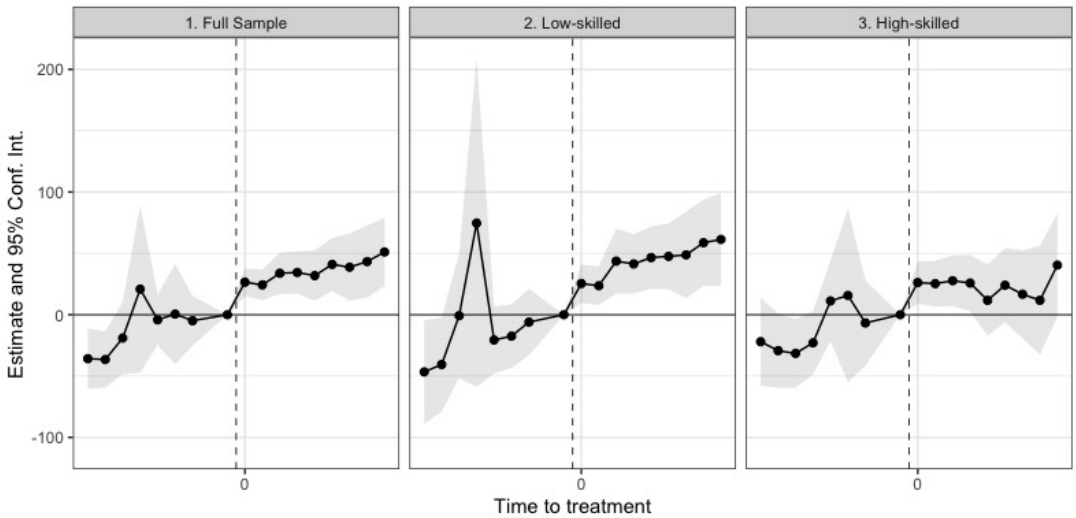
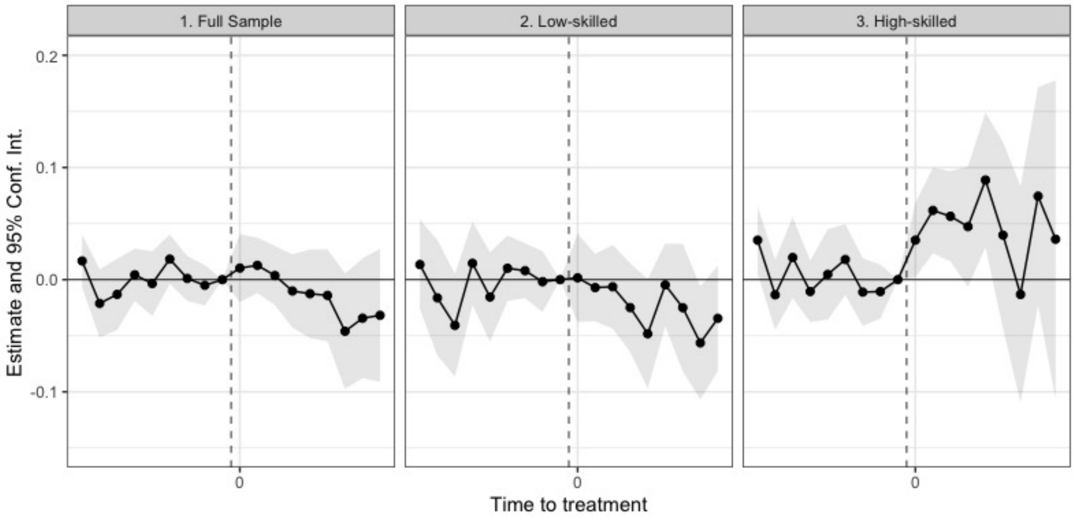
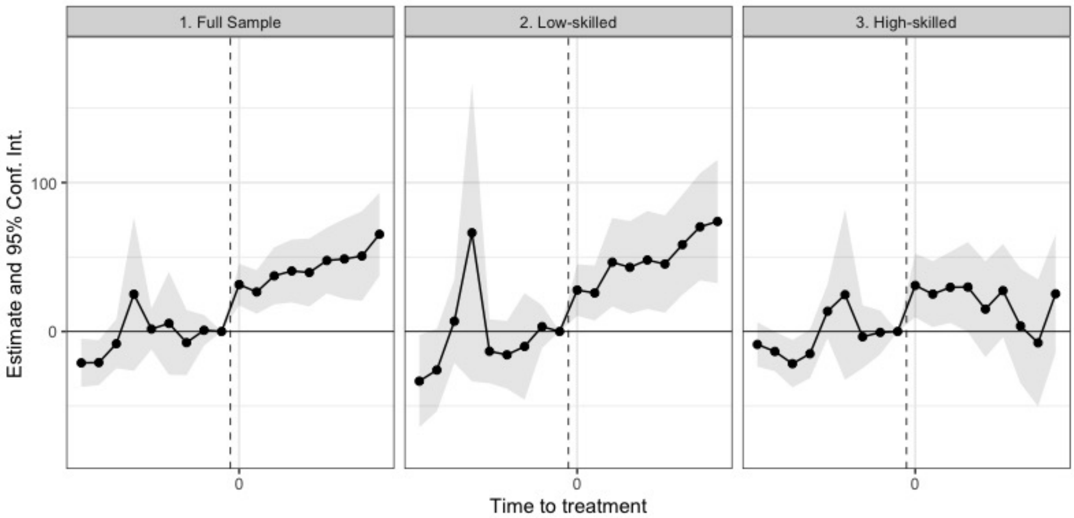


Figure notes: Event Study Analysis based on Sun and Abraham (2021) showing treatment effects on $y_{i,t}^{QUAN}$ for each relative treatment period. The sample includes the time between October 2010 and December 2011. The estimates use *t-1* and *t-2* as reference category. $y_{i,t}^{QUAN}$ is the work speed outcome expressed in the agents average handling time.

Figure A7: Two-Stage DiD Estimates based on Gardner (2021)



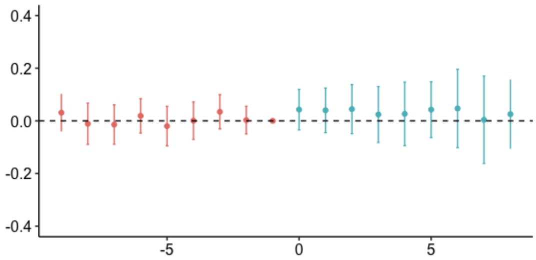
(a) Service Quality - y_{it}^{QUAL}



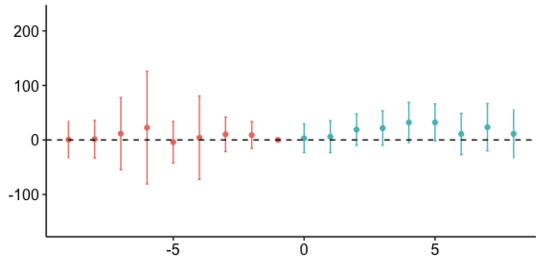
(b) Work Speed - y_{it}^{QUAN}

Figure notes: Event Study Analysis based on Gardner (2021) showing treatment effects on $y_{i,t}^{QUAL}$ (a) and $y_{i,t}^{QUAN}$ (b) for each relative treatment period and dependent on the level of skill. The sample includes the time between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the service quality outcome represented by the agent's average handling time.

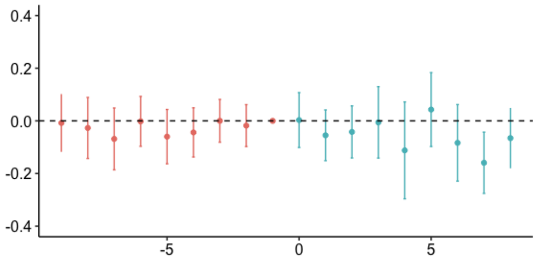
Figure A8: Estimates based on Callaway and Sant'Anna (2021)



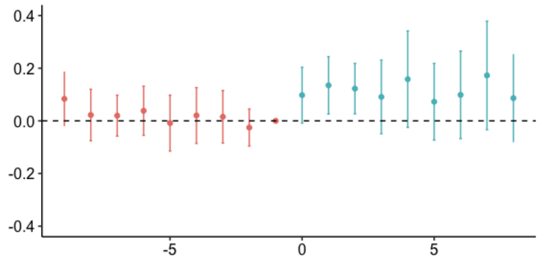
(a) y_{it}^{QUAL} - Full Sample



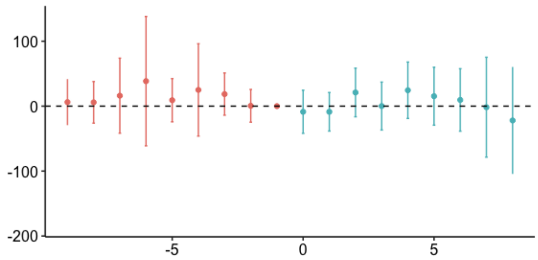
(b) y_{it}^{QUAN} - Full Sample



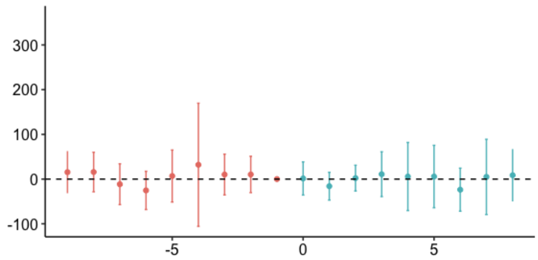
(c) y_{it}^{QUAL} - Low-skilled



(d) y_{it}^{QUAL} - High-skilled



(e) y_{it}^{QUAN} - Low-skilled



(f) y_{it}^{QUAN} - High-skilled

Figure notes: Event Study Analysis based on Callaway and Sant'Anna (2021) showing treatment effects on $y_{i,t}^{QUAL}$ (a and b) and $y_{i,t}^{QUAN}$ (c and d) for each relative treatment period and dependent on the level of skill. The sample includes the time between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the service quality outcome represented by the agent's average handling time.

Table A1: The Effect of the Bonus Introduction - Skill Differences (2)

Dependent Variables:	Service Quality (y_{it}^{QUAL})			Work Speed (y_{it}^{QUAN})		
Skill-level	Low	Medium	High	Low	Medium	High
	(1)	(2)	(3)	(4)	(5)	(6)
Performance Pay	-0.052**	0.025	0.072*	24.3*	36.8**	13.4
	(0.024)	(0.026)	(0.039)	(12.5)	(16.1)	(14.6)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,057	1,378	1,082	1,057	1,378	1,082
R^2	0.428	0.324	0.302	0.705	0.564	0.751
Mean Dependent variable	0.242	0.419	0.534	337.7	345.1	343.7
N Agents	130	129	129	130	129	129

Table notes: The results display average treatment effects of the bonus introduction on $y_{i,t}^{QUAL}$ and $y_{i,t}^{QUAN}$ depending on the level of skill. Treatment effects for low-skilled agents are shown in Columns (1) and (4), medium-skilled agents in Columns (2) and (5), and high-skilled agents in Columns (3) and (6). The sample includes the months between October 2010 and December-2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A2: The Effect of the Bonus Introduction on Service Quality - Disaggregated Effects

Dep. Var.: Share of...	Bad Calls ($N_{it,0-6}$)			Neutral Calls ($N_{it,7-8}$)			Good Calls ($N_{it,9-10}$)		
Sample	Full	Low-skilled	High-skilled	Full	Low-skilled	High-skilled	Full	Low-skilled	High-skilled
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Performance Pay	0.015	0.044	-0.048	-0.016	-0.008	-0.018	0.0003	-0.036	0.065**
	(0.022)	(0.029)	(0.031)	(0.019)	(0.027)	(0.029)	(0.019)	(0.024)	(0.030)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,517	1,777	1,740	3,517	1,777	1,740	3,517	1,777	1,740
R^2	0.458	0.502	0.440	0.387	0.322	0.459	0.326	0.370	0.353
Mean Dependent Variable	0.41	0.54	0.29	0.37	0.33	0.42	0.21	0.13	0.29
N Agents	388	194	194	388	194	194	388	194	194

Table notes: The results display average treatment effects of the bonus introduction on different components of $y_{i,t}^{QUAL}$ depending on the level of skill. The sample includes the months between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * p < 0.1, ** p < 0.05, *** p < 0.01.

Table A3: Baseline Estimates With Adjusted Reference Category

Dependent Variables:	Service Quality (y_{it}^{QUAL})			Work Speed (y_{it}^{QUAN})		
Sample	Full (1)	Low-skilled (2)	High-skilled (3)	Full (4)	Low-skilled (5)	High-skilled (6)
Performance Pay	0.004 (0.015)	-0.024 (0.017)	0.062*** (0.024)	34.3*** (8.26)	40.9*** (10.8)	23.4** (11.7)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,517	1,777	1,740	3,517	1,777	1,740
R^2	0.299	0.369	0.287	0.598	0.600	0.694
Mean Dependent variable	0.398	0.295	0.501	342.2	333.4	350.9
N Agents	388	194	194	388	194	194

Table notes: The results display average treatment effects of the bonus introduction on $y_{i,t}^{QUAL}$ and $y_{i,t}^{QUAN}$ depending on the level of skill. Treatment effects for low-skilled agents are shown in Columns (1) and (3) and high-skilled agents in Columns (2) and (4). The estimates use *t-1*, *t-2* and the first period as reference category. The sample includes the months between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Estimates Based on Gardner (2022) and Callaway and Sant'Anna (2021)

Dependent Variables:	Service Quality (y_{it}^{QUAL})			Work Speed (y_{it}^{QUAN})		
Sample	Full (1)	Low-skilled (2)	High-skilled (3)	Full (4)	Low-skilled (5)	High-skilled (6)
Specification						
Performance Pay (Gardner, 2022)	-0.009 (0.0115)	-0.0189 (0.0124)	0.0488** (0.0191)	40.63*** (8.790)	44.45*** (12.06)	21.69* (12.55)
Performance Pay (Callaway and Sant'Anna, 2021)	0.034 (0.032)	-0.046* (0.024)	0.115* (0.037)	17.476* (8.927)	20.051 (12.494)	-0.039 (13.933)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,517	1,777	1,740	3,517	1,777	1,740
Mean Dependent Variable	0.398	0.295	0.501	342.2	333.4	350.9
N Agents	388	194	194	388	194	194

Table notes: The results display average treatment effects of the bonus introduction on $y_{i,t}^{QUAL}$ and $y_{i,t}^{QUAN}$ depending on the level of skill. Treatment effects for low-skilled agents are shown in Columns (1) and (3) and high-skilled agents in Columns (2) and (4). The sample includes the months between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: The Effect of the Bonus Introduction - Peer Effects

Dependent Variables:	Service Quality (y_{it}^{QUAL})			Work Speed (y_{it}^{QUAN})		
Sample	Full (1)	Low-skilled (2)	High-skilled (3)	Full (4)	Low-skilled (5)	High-skilled (6)
Panel A						
Performance Pay	-0.007	-0.047**	0.075***	17.3**	20.6*	11.8
	(0.019)	(0.023)	(0.029)	(7.41)	(10.5)	(11.7)
Team-FE	X	X	X	X	X	X
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,498	1,761	1,737	3,498	1,761	1,737
R^2	0.316	0.392	0.325	0.625	0.643	0.729
Mean Dependent Variable	0.399	0.301	0.498	344.9	338.7	351.1
N Agents	379	190	189	379	190	189
Panel B						
Performance Pay	-0.002	-0.031	0.065**	19.3***	20.2**	11.6
	(0.018)	(0.023)	(0.029)	(6.78)	(8.05)	(11.5)
$\overline{T_{post}}$	0.039	0.068	-0.020	-31.5**	-20.2	-46.5**
	(0.052)	(0.077)	(0.070)	(14.9)	(20.4)	(19.7)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,920	1,478	1,442	2,920	1,478	1,442
R^2	0.324	0.401	0.340	0.636	0.634	0.724
Mean Dependent Variable	0.405	0.305	0.501	344.40	334.71	353.77
N Agents	355	175	180	355	175	180

Table notes: The results display average treatment effects of the bonus introduction on $y_{i,t}^{QUAL}$ and $y_{i,t}^{QUAN}$ depending on the level of skill. Treatment effects for low-skilled agents are shown in Columns (1) and (3) and high-skilled agents in Columns (2) and (4). The sample includes the months between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: The Effect of the Bonus Introduction on Evaluated Calls

Dep. Var.	Evaluated calls		
	Full sample (1)	Low-skilled (2)	High-skilled (3)
Performance Pay	4.05*	3.69	5.80*
	(2.30)	(3.34)	(3.15)
Control variables	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	3,517	1,777	1,740
R^2	0.754	0.781	0.757
Mean Dependent Variable	11.85	11.90	11.79
N Agents	388	194	194

Table notes: The results display average treatment effects of the bonus introduction on the number of evaluated calls, the share of evaluated calls and the standard deviation of evaluated calls depending on the level of skill. The sample includes the months between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Alternative Specification - Broader Skill Measure

	Service Quality (y_{it}^{QUAL})			Work Speed (y_{it}^{QUAN})		
Sample	Full (1)	Low-skilled (2)	High-skilled (3)	Full (4)	Low-skilled (5)	High-skilled (6)
Performance Pay	-0.007	-0.039	0.047*	25.5***	23.6**	23.8**
	(0.018)	(0.024)	(0.027)	(7.64)	(9.65)	(11.4)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,697	1,825	1,872	3,697	1,825	1,872
R^2	0.331	0.363	0.299	0.615	0.639	0.687
Mean Dependent Variable	0.39	0.28	0.50	333.82	330.82	336.82
N Agents	484	242	242	484	242	242

Table notes: The results display average treatment effects of the bonus introduction on $y_{i,t}^{QUAL}$ and $y_{i,t}^{QUAN}$ depending on an alternative skill measure. The sample includes the months between October 2010 and December 2011. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Alternative Specification - More Balanced Sample

	Service Quality (y_{it}^{QUAL})			Work Speed (y_{it}^{QUAN})		
Sample	Full	Low-skilled	High-skilled	Full	Low-skilled	High-skilled
	(1)	(2)	(3)	(4)	(5)	(6)
Performance Pay	-0.007	-0.038	0.053*	25.3***	33.1***	14.7
	(0.019)	(0.024)	(0.028)	(7.81)	(10.3)	(11.0)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,948	1,502	1,446	2,948	1,502	1,446
R^2	0.299	0.367	0.317	0.548	0.565	0.659
Mean Dependent Variable	0.40	0.32	0.39	326.94	320.94	333.18
N Agents	249	127	122	249	127	122

Table notes: The results display average treatment effects of the bonus introduction on $y_{i,t}^{QUAL}$ and $y_{i,t}^{QUAN}$ depending on the level of skill. The sample includes the months between and conditions on being observed in the period before and after the bonus was introduced. $y_{i,t}^{QUAL}$ is the service quality outcome expressed in the agent's net promoter score (NPS). $y_{i,t}^{QUAN}$ is the work speed outcome represented by the agent's average handling time. The mean of the dependent variable refers to the last observed period before the bonus was introduced for the first time. Standard errors are clustered at the agent level. * p < 0.1, ** p < 0.05, *** p < 0.01.