WORKING PAPER 2025:6

Making the right call: the heterogeneous effects of individual performance pay on productivity

Marco Clemens
Jan Sauermann

IFAU
INSTITUTE FOR EVALUATION OF LABOUR MARKET AND EDUCATION POLICY

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala.

IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

Making the right call: the heterogeneous effects of individual performance pay on productivity[a]

Marco Clemens[b] Jan Sauermann[c]

February 27, 2025

Abstract:

Performance pay has been shown to have important implications for worker and firm productivity. Although workers' skills may directly matter for the cost of effort to reach performance goals, surprisingly little is know about the heterogeneity in the effects of incentive pay across workers. In this study, we apply a dynamic difference-in-differences estimator to the introduction of a generous bonus pay program to study how salient performance thresholds affect incentivized and non-incentivized performance outcomes for low- and high-skilled workers. While we do find that individual incentive pay did not affect workers' performance on average, we show that this result conceals an underlying heterogeneity in the response to individual performance pay: individual performance pay has a significant effect on the performance of high-skilled workers but not for low-skilled workers. The findings can be rationalized with the idea that the costs of effort differ by workers' skill level. We also explore whether agents alter their overtime hours and find a negative effect, possibly avoiding negative consequences of longer working hours.

Keywords: performance pay, incentives, productivity, skills, panel data

JEL-codes: M52, J33, C23

[a]We thank Spencer Bastani, Andries De Grip, Lazlo Goerke, Colin Green, Patrick Kampkötter, Olof Rosenqvist, Timo Vogelsang, two anonymous reviewers, and seminar and conference participants at the RES PhD Conference 2023 (Glasgow), IAAE 2023 (Oslo), WLE 2023 (Trier), COPE 2024 (Zurich), and two anonymous reviewers of the HWS 2023 (Berlin) for valuable comments and feedback. Declarations of interests: none.

[b]*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

[c]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 working from home will likely make performance-related pay more prevalent in the future (Eurofund, 2020; Barrero et al., 2023). At the same time, performance-related 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-related pay has shown that introducing worker-level monetary incentives 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., 2023), but also that the effects differ substantially between studies (Havranek et al., 2022). This heterogeneity between studies can be caused by differences in incentive schemes or settings, e.g. due to the size of the monetary incentive (Gneezy and Rustichini, 2000) or whether the outcome ise qualitative or quantitative. Studies have also provided evidence for substantial heterogeneity *within* firms if workers differ in risk-preferences (Cadsby et al., 2007; Grund and Sliwka, 2010), skills (Kowalski, 2019), or other worker characteristics (Opitz et al., 2024).

In this study, we provide evidence for one channel through which worker characteristics can be crucial for the effectiveness of individual performance pay in firms. We examine the introduction of individual performance bonuses for employees in a call center operated by a multinational telephone company in the Netherlands to address the research questions of whether and how these bonuses influence worker performance and whether this effect differs between low- and high-skilled workers. We employ dynamic difference-in-difference methods (Sun and Abraham, 2021) to compare workers eligible for individual 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 individual performance pay programs when performance thresholds are defined. Compared to low-skilled individuals, high-skilled individuals have a lower cost of effort to reach performance thresholds. In line with this argument, we document that individual performance bonuses have the expected effect for high-skilled workers, whereas low-skilled individuals do not react to its introduction. 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. We find that high-skilled workers' non-incentivized outcome is not affected by the introduction of the new incentive system whereas low-skilled individuals significantly *increase* the time spent on each call. These results show that the introduction of this program led to advantages for high-skilled agents on both outcomes, relative to low-skilled agents. Third, we also estimate effects of the performance pay program on working hours. Workers can have an incentive to work extra hours to achieve higher bonuses. In contrast to this, however, we find that workers are less likely to work overtime. We argue that this can be explained by workers trying to limit the risk of potentially negative evaluations which might be more likely when working longer hours.

This paper contributes to the literature in several ways. First, we aim at contributing to the the literature documenting the 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, and find rather inconclusive results. 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 individual 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. For managers, this highlights the importance of calibrating performance thresholds appropriately. For the academic literature, this result can potentially contribute to explaining sorting effects of performance pay.

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, 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 no 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. This suggests that agents either shift effort from the non-incentivized measure in an unsuccessful attempt to reach performance thresholds or that they simply shirk on non-incentivized outcomes.

Third, we add to the scarce literature on the effects of performance pay on *quality*-related measures of performance. Compared to quantitative performance outcomes, it is more difficult for firms to monitor the quality of workers' output. Only a handful of

[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.

studies use quality-related outcomes and they are often assessed as the non-incentivized outcome. Asch (1990) uses information on the number, but also on the education of recruited soldiers. Kato and Shu (2008) use data on total output and information on the defect rate of textile workers in a Chinese weaving company. Shi (2010) uses information on performance quality, as well as performance quantity of workers thinning fruit trees. If quality (or a creative task) is directly incentivized, recent evidence mostly comes from the lab with mixed results: For instance, Bradler et al. (2019) conducted an experiment with more than 1,000 participants to incentivize the number and quality of creative ideas and discovered an incentive effect for both outcomes. Kleine (2021), on the other hand, ran an experiment with 460 participants, incentivizing the number of problems solved. According to the findings, financial incentives can 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 their call with an individual agent.

Last, 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, 2024; 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. In Section 3, we present main outcome variables and our definition of skill. Section 4 contains the methodological framework, and Section 5 the main results. We offer robustness checks in Section 6, and conclude in Section 7.

2 The firm and its incentive schemes

In this section, we 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] The call center is an in-house service center that handles the inbound calls of current and prospective customers and is organized into different departments that cover different customer and call types. Customers can call in when they have questions, problems or complaints and are connected through an automated routing system to available agents.

The main task of call agents is to handle these inbound customer calls. These can be of administrative or technical nature and can range from rather simple calls, such as simple technical questions, to more complex issues, such as problems with billing. In addition to answering calls and talking to customers, agents are also required to access and enter information in the customer database during and after each call. 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

The firm changed its performance incentives from the old system, in which an annual performance appraisal determined bonus and wage growth to a new system, in which salient short-term performance targets were defined for agents.[3]

Under the old system, which was in place up until March 2011, each agent had annual appraisal interviews with the respective team leader, typically held in April and May. These interviews resulted in five grades 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,

[2]Data from the same firm are also used in, e.g., De Grip et al. (2016) and Sauermann (2023). Our paper uses a different and so far entirely unused sample that covers the introduction of performance bonuses and more departments than were previously analyzed.

[3]Implicit incentives, such as promotion incentives, are arguably rather small. While team leader positions are often filled with prior call agents, there are only relatively few positions.

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.

The new incentive system, which was introduced for all agents in April 2011, differed in two ways from the old system.[5] First, the annual bonus was replaced by a new, short-term incentive pay with explicitly defined performance thresholds that were salient to the agents. Only one performance outcome, the net-promoter score (NPS), was used to calculate the bonus payments. This was done to increase satisfaction and thereby improve competitiveness against other firms.[6] For each evaluation period, which consisted of three months, management set NPS-performance thresholds and communicated those to all agents. At any time during the evaluation period, agents could observe their individual performance so far. In the month following the evaluation period, agents received the bonus in addition to their regular salary. Second, some groups of agents were exempted from the individual performance pay and instead received a bonus based on the performance of their department. These agents are employees during their probationary period, employees on contracts from temporary help agencies, and agents with relatively short working hours who receive less than 60 evaluated calls per evaluation period. Also for these agents, both performance thresholds and departmental performance were made

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

[6]Indeed, recent meta-evidence (Otto et al., 2020) suggests a positive relationship between customer satisfaction and firm-level performance outcomes such as profits and hence competitiveness.

salient. Management adjusted the level of the performance thresholds for the following evaluation period in the last week of the preceding evaluation period.

Figure A1 shows how performance translates into bonuses. There are five bonus levels $B_1, \dots, B_5$, which correspond to bonuses of 0, 4.8, 8, 10, and 12% of the wages earned during the quarter, respectively. 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.[7]

3 Data

The data used in this study contain monthly performance measures of each individual agent nine months before, and nine months after the introduction of the bonus related to service quality in April 2011. That is, we observe performance starting in July 2010 and ending in December 2011. In this section, we introduce the main outcome variables service quality and work speed, as well as our definition of workers' skills.

3.1 Outcome variables

Call centers typically have several performance measures covering different dimensions of an agents' productivity. The call center management of the firm we observe in this study predominantly focuses on two measures of performance: the NPS, which measures *service quality* provided by an agent, and average handling time, which measures how fast agents provide their service (*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.

[7]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.

Service quality The measure of service quality is on NPS, which is based on customer satisfaction surveys among a randomly chosen population of customers. The NPS measure is very widespread when evaluating customer agents' performance and is reported to be used by more than two-thirds of Fortune 1000 (Baehre et al., 2022). The measure itself is defined such that it is correlated to customer loyalty (see, e.g., Keiningham et al., 2007).

Among other questions, customers were asked how likely they would recommend the mobile operator to family and friends, based on the previous call with agent i.[8] The answers can range from 0 ('very unlikely') to 10 ('very likely'). Based on this, for each agent, calls are classified into three categories: "Detractors," for scores less than 7, "Passives," for scores from 7 or 8, and "Promoters," for scores of 9 or 10.[9] The NPS is calculated by taking the difference between the percentages of Promoters and Detractors, and then expressing this difference as a proportion of the total number of respondents. Formally, the NPS can be expressed as

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

where $N_{it,9-10}$ denotes the number of Promoters indicating that customers evaluate the agent's service as outstanding (9 or 10). $N_{it,0-6}$ denotes the number of customers who rate the agent's service less favorably (Detractors) and $N_{it,0-10}$ is the overall number of NPS evaluations. As such, $y_{i,t}^{QUAL}$ is defined on a scale from -1 to +1 with higher values related to higher quality provided by agents. 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.

To be able to use NPS as a reliable measure of service quality, we see two potential concerns. First, if agents would know whether the call will be evaluated, agents might be biased in their behavior towards the customer. Importantly, however, the agent has no way to see whether (1) the customer agreed to be called back, and (2) the customer

[8]The exact question was 'Based on this contact, how likely are you to recommend [the firm] to your family and your friends?'.

[9]Before the bonus was introduced, the percentage of "Promoters" in our sample was 21.3%, the percentage of "Passives" was 35.0%, and the percentage of "Detractors" was 43.6%. These shares marked at 22.3%, 39.4%, and 38.3%, respectively, in the period after the introduction of the bonus.

will in fact be called back. Second, customers who received service of lower or higher quality could be more or less likely to participate in the customer survey even after having initially agreed to participate in the survey. While it is difficult to rule out such effects, it is not obvious in which direction this should bias the results and, importantly, that it would be related to 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 (AHT) per call 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

The main aim of this paper is to show whether skilled workers react differently to performance incentives. In the absence of a direct measure of workers' skill, previous studies, such as Lazear (2000) and Kowalski (2019), have used prior output as measures of skills with the idea that these are correlated with the underlying ability and not directly correlated with current output.

Based on this procedure, we construct a measure of skill relying on performance measures before the introduction of the new performance pay system. For this purpose, we use the two main performance outcomes used in the call center to assess the performance of workers: service quality and work speed.[10] Relying only on data from the six months

[10] We base our skill measure on both incentivized and unincentivized outcomes to allow for the possibility that agents could trade greater quality for lower work speed, and vice versa. The premise is that the same level of quality obtained with a shorter average handling translates into higher skill levels.

before the bonus introduction, we first residualize each performance outcome, create rankings for both outcomes and then create an overall index combining the agents' position in both rankings. Based on this index, we define workers as high-skilled and low-skilled when they are ranked above the median or ranked below the median, respectively.

Specifically, we first use data on both outcome measures from October 2010 to March 2011, to estimate the following regression equation:

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

With $y^j_{i,t}$ either being service quality ($y^{QUAL}_{i,t}$) or work speed ($y^{QUAN}_{i,t}$). X'_{it} comprises 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. This allows to create a ranking that abstracts from other time or individual-specific variation.

For each agent-month observation, we then extract the residuals $\widehat{\varepsilon}_{i,t}$, indicating service quality and work speed differences that are unexplained by the aforementioned control variables. Next, for each outcome, we rank the 388 observed agents based on these residuals, i.e. comparing skill levels in terms of service quality and work speed across all agents. To define a combined measure, we use the sum of both rankings as a ranking score. For instance, an agent that performs well in service quality (rank score 388) and badly in work speed (rank score 1) will end up in the middle of the final skill distribution. Figure A2 illustrates that the ranking score distribution is relatively normal. This suggests that some agents do poorly in both service quality and work speed, while others excel in both. With that distribution, we define agents as high-skilled if the total individual ranking score is greater than the median ranking score of all agents, and as low-skilled otherwise.

While it is impossible to assess how well our skill measure relates to an agent's actual skill level, we contrast it with customer assessments that are likely correlated with an agent's *actual* skill level. In customer satisfaction surveys, customers are asked to assess whether or not the problem was solved, the agent's effort level, the agent's expertise,

Table 1: Correlation of Skill with other Service Quality Measures

Dependent Variables:	Problem Solved (1)	Effort (2)	Expertise (3)	Overall Rating (4)
High-skilled	0.317***	0.327***	0.363***	0.303***
	(0.074)	(0.074)	(0.071)	(0.070)
Constant	-0.177***	-0.236***	-0.266***	-0.241***
	(0.052)	(0.052)	(0.050)	(0.050)
Observations	928	928	928	928
R^2	0.01956	0.02059	0.02736	0.01970
Adjusted R^2	0.01850	0.01953	0.02631	0.01864

Notes: The table displays the correlation of being classified as high-skilled (0/1) with customer satisfaction measures in the two quarters leading up to the bonus introduction (October 2010 - March 2011).

and an overall rating of the agent's performance.[11] Table 1 shows correlation coefficients between our skill measure and customer assessments. All customer assessments are standardized with zero mean and a standard deviation of one. The table shows that for all four customer assessments, being defined as high-skilled is related to higher ratings.

In Section 5.2 and 6, we provide additional robustness tests regarding the skill definition. First, we extend that specification using a quartile sample split instead of splitting at the median index. Second, we also provide an alternative approach using the full pre-bonus period. Last, we use a simple ranking instead of a regression-based skill definition based on residuals discussed in this section.

4 Methodology

4.1 Estimation strategy

A conventional way to estimate the treatment effects of a policy intervention is using dynamic two-way fixed effects (TWFE) estimators of the following form:

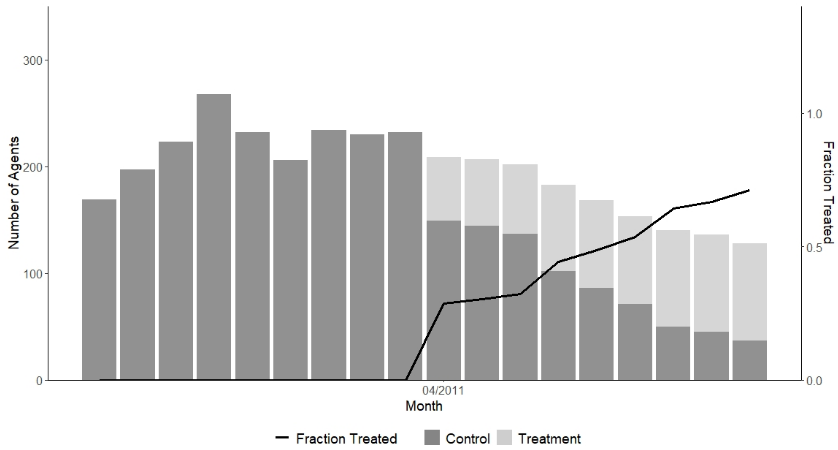
$$y_{it}^{j} = \delta_i + \lambda_t + \alpha_1 X_{it}' + \sum_{l=-K}^{-2} \beta_l D_{it}^{l} + \sum_{l=0}^{L} \beta_l D_{it}^{l} + \epsilon_{it} \tag{3}$$

[11] The exact phrases are "Was your problem solved by this call?" and "How would you rate the call agent's effort/expertise/overall performance on a scale from 1 (poor) to 5 (great)?". These measures stem from the same customer satisfaction survey as the NPS measure described in section 3.1.

where the outcome y_{it}^{j} is the performance measure j of agent i in month t, δ_i accounts for individual fixed effects, and λ_t captures time-fixed effects.[12] D_{it} defines the treatment status of agent i. The estimate of interest, β_l can be used to assess both the parallel trends assumption in the periods l before the introduction of the new performance pay system and the periods after to assess the treatment effect of the introduction of the performance pay. X'_{it} comprises a vector of control variables including the number of planned working hours, tenure, sub-department dummies, the number of evaluated calls, and a dummy for being a temporary help agent.

In our setting, however, treatment is defined as an individual agent's eligibility to participate in the individual performance incentive scheme in a given month. Agents become eligible for the individual incentive pay when transitioning from temporary help agent status to regular contracts with the call center, or when exceeding the six-month tenure threshold. While many agents become eligible for individual performance pay with its introduction, others become eligible only later. Figure 1 displays the share of agents in treatment and control groups. The share of treated agents increases from approx. 30% in April 2011 to around 70% in the last sample month.

Figure 1: Distribution of Treatment and Control Groups over Time



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.

[12]Individual fixed effects require within agent variation in the outcomes. In our data, the between agent variation, indicated by the mean squares, for $y_{i,t}^{QUAL}$ is approximately three times higher than the within agent variation. For $y_{i,t}^{QUAN}$ this ratio increases to 6. However, this indicates that while between-agent variation is dominant, within-agent variation is still relevant.

This implies that the new performance pay system was introduced under staggered adoption. In such cases, it has been shown that dynamic TWFE may produce biased estimates of the true treatment effect (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Roth et al., 2023). 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., 2023). Hence, to overcome this limitation, we estimate both TWFE treatment effects and treatment effects following the method proposed by Sun and Abraham (2021), which allows 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 individual bonus as a control group, to allow valid comparisons. With that, we obtain treatment effects based on the weighted average of coefficients for each cohort and each relative time after or before the treatment.[13]

In our setting D_{it}^{l} represents the relative time period l of agent i with respect to the first time the individual bonus was applied ($l = 0$). Furthermore, we include variables that may be related to both eligibility for the new bonus and service outcomes. Because service quality is derived from customer surveys, we also control for the number of customer evaluations. Hence, β_l, with $l > 0$, then indicate the coefficients of interest and the effect of the individual bonus introduction on performance, conditioned on covariates, common time-trends, and individual time-invariant factors.

4.2 Interpreting treatment effects

As described in Section 2.2, all agents had the same incentives prior to the introduction of the new incentive scheme: an annual appraisal interview without explicit targets determined the (annual) bonus. From April 2011 onward, management set explicit performance thresholds that were clearly communicated to all agents and paid out after every third month. Agents in the treatment group received performance bonuses based

[13]For a detailed overview of the method and its assumptions we refer to Sun and Abraham (2021) and Roth et al. (2023).

on their own performance relative to the performance thresholds; agents in the control group received performance bonuses based on their department's average performance relative to the performance thresholds.

This implies that the counterfactual to the treatment (receiving individual performance incentives) is a weighted mix of annual performance bonuses (before April 2011) and salient group incentives (from April 2011 onward). Because both counterfactuals may have some incentive effects that are plausibly smaller than the incentive effects of the individual incentive pay treatment, we argue that the treatment effects estimated in this paper serve as a lower bound of the counterfactual of having no incentives in place.[14] Furthermore, it is important to stress that we are primarily interested in documenting differences in treatment effects *across* group, i.e. between low-skilled and high-skilled workers.

4.3 Estimation sample and descriptive statistics

Because we define individuals as either high-skilled or low-skilled based on their performance in the six months before the introduction of the new incentive scheme, we include all agents who worked at least one month during this period. The overall estimation sample includes all observations of these agents between July 2010 and December 2011.[15] The total number of agents in the estimation sample is 388 with 3,517 agent-month observations.

Table 2 shows differences in characteristics between eventually treated and never-treated individuals in the last observed month before the bonus system was introduced (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 2 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

[14]To account for potential differences between departments with different size (and thus different group incentive effects), we also tested controlling for department size. This does not change our results.
[15]Potential issues with the sample restriction are addressed in Section 6.

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.

Table 2: 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			

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.

Looking at service quality ($y_{i,t}^{QUAL}$) and work speed ($y_{i,t}^{QUAN}$), there is no difference in service quality between treated and untreated agents, but calls are 23 seconds shorter for treated agents, which translates into a *better* performance among agents who are eventually treated. This is in line with earlier research showing that agents with longer tenure have on average higher performance (De Grip et al., 2016). Throughout the paper, we distinguish between high-skilled and low-skilled employees. Figures A3 and A4 in the Appendix further illustrate the development of service quality and quantity over time by skill level. 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 2). 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.

5 Results

5.1 Main results

Table 3 shows the treatment effect of the new individual incentive scheme on the incentivized measure, service quality $y_{i,t}^{QUAL}$ (Columns (1)-(4)), and the non-incentivized measure, work speed $y_{i,t}^{QUAN}$ (Columns (5)-(8)), respectively. For both outcomes, we report TWFE estimates (Columns (1), (2), (5) and (6)) and estimates based on Sun and Abraham (2021) (Columns (3), (4), (7) and (8)). All regressions control for time (month) and worker fixed effects.

With or without including control variables, the results show no significant treatment effects in any of the specifications for $y_{i,t}^{QUAL}$ despite agents in the treatment group receiving well-defined monetary bonuses with salient performance thresholds. The implementation of the new individual performance incentives had no measurable effect on service quality.

Table 3: The Effect of the Bonus Introduction on Service Quality and Work Speed

Dep. Variables	Service Quality (y_{it}^{QUAL})				Work Speed (y_{it}^{QUAN})			
	TWFE		Sun and Abraham (2021)		TWFE		Sun and Abraham (2021)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Performance Pay	0.002	-0.0007	0.000	-0.008	30.6***	36.4***	18.5***	25.5***
	(0.011)	(0.011)	(0.017)	(0.018)	(8.06)	(8.03)	(6.91)	(7.70)
Hours planned		0.0003		0.0001		0.359		0.473
		(0.0008)		(0.0008)		(0.463)		(0.466)
Tenure		0.008		0.009		-4.54		-2.74
		(0.010)		(0.010)		(6.22)		(6.05)
# Eval. Calls		0.000		0.000		-0.728***		-0.799***
		(0.0002)		(0.0002)		(0.152)		(0.165)
Temp. Agent		-0.005		-0.002		22.4***		18.6***
		(0.011)		(0.012)		(7.44)		(7.09)
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	0.281	0.299	0.529	0.567	0.564	0.598
Mean dep. variable			0.405				333.24	
N Agents				388				

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 subsequentially 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 show a positive treatment effect, i.e. agents require longer time per call. There are two main explanations for this result: first, with the introduction of the new incentive system agents could shift their effort from non-incentivized work speed to incentivized service quality. An alternative, related, explanation could be that agents take more time to improve on the incentivized performance measure, i.e., service quality. We will further explore this in Subsection 5.2.

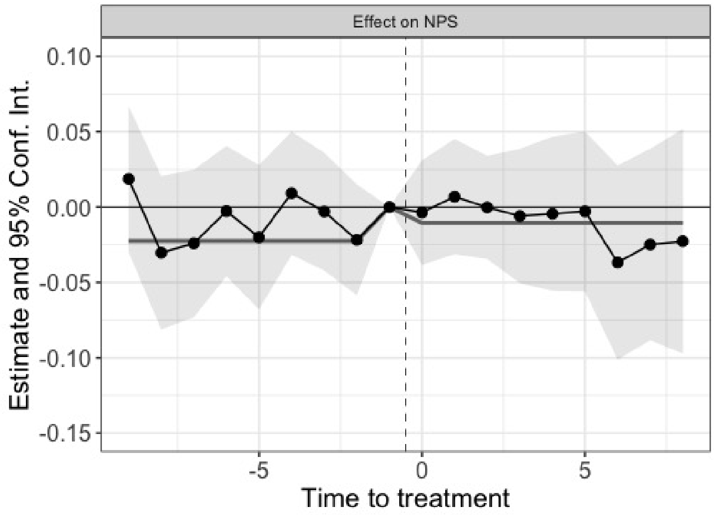
The results also show that the point estimates are slightly larger when accounting for control variables in both the TWFE and estimates based on Sun and Abraham (2021). One reason for that result may be the fact that agents become eligible for treatment only if they are not employed on a temporary basis 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 any longer. Therefore, the initial estimates may *underestimate* the (negative) treatment effect in terms of work speed.[16] 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 setups. 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 3).

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 preperiods. For the work speed measure y_{it}^{QUAN} (Sub-figure (b)), some periods seem to violate the parallel trends assumptions. Therefore, we use alternative event-study de-

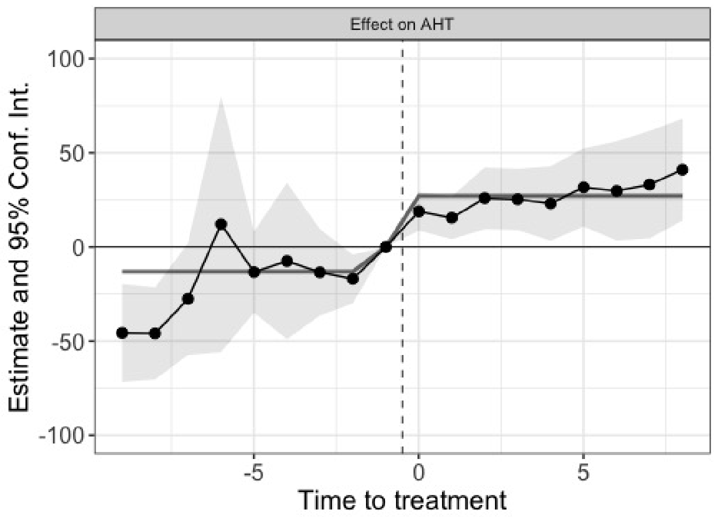
[16] An alternative channel could be higher employment protection legislation due to permanent contracts leading to lower productivity levels (Ichino and Riphahn, 2005; Mainar et al., 2018). If we do not control for being a temporary help agent, this should rather lead to an overestimation of the (negative) treatment effect on work speed. Table A1 shows that when excluding agents entering the treatment group at the same time as they move from being a temporary help agent to a regular contract, the estimated treatment effects without control variables for work speed (columns 5 and 7) are rather amplified compared to Table 3. This adds evidence to our initial view that coefficients are underestimated.

signs and model specifications in Section 6 to further investigate potential violations of the parallel trends assumption.

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})

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. The grey line represents the average treatment effect in the pre and post-treatment periods.

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 4, 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 that low-skilled agents do not significantly react to the introduction of individual performance incentives. In fact, the point estimate is even negative. High-skilled individuals, however, have a positive treatment effect on y_{it}^{QUAL}. 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 5.1 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% (0.051/0.448). This effect is statistically larger than the point estimate for low-skilled individuals.

For the non-incentivized outcome y_{it}^{QUAN}, which is based on average handling time, Table 4 reveals that the overall effects on work speed are driven by relatively low-skilled agents. Given the pre-introduction mean of approximately 350 seconds, the bonus introduction led to a decrease in y_{it}^{QUAN} by 11% (37.7/350). The coefficient for high-skilled agents remains statistically insignificant and close to zero.

The results suggest that high-skilled agents can react to individual incentives by providing higher performance while maintaining performance levels on other dimensions. Low-skilled agents, however, reduce effort in the performance dimension that is not incentivized anymore but at the same time fail to react to the performance incentive. The estimates could also suggest that low-skilled agents deliberately take more time to provide greater quality but eventually fail to translate the additional time into higher incentivized outcomes. They may not know how to effectively trade between work speed and service quality. These effects are estimated with worker-fixed effects suggesting that this is not due to individual-specific time-invariant characteristics.

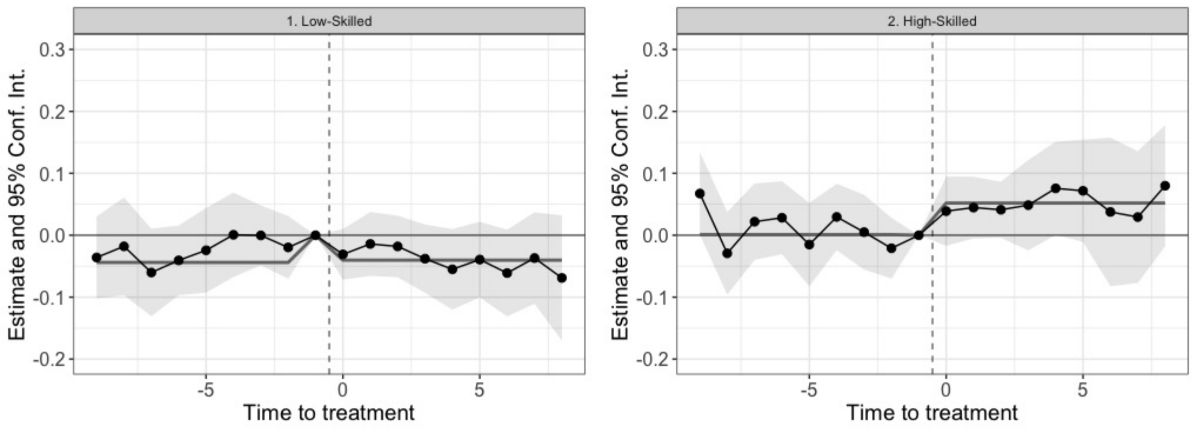
Table 4: The Effect of the Bonus Introduction - Skill Differences

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	-0.037	0.051**	37.7***	1.31
	(0.023)	(0.026)	(10.1)	(8.65)
Control variables	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	1,681	1,836	1,681	1,836
R^2	0.347	0.306	0.647	0.752
Mean dependent Variable	0.363	0.448	349.576	316.620
N Agents	194	194	194	194

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.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

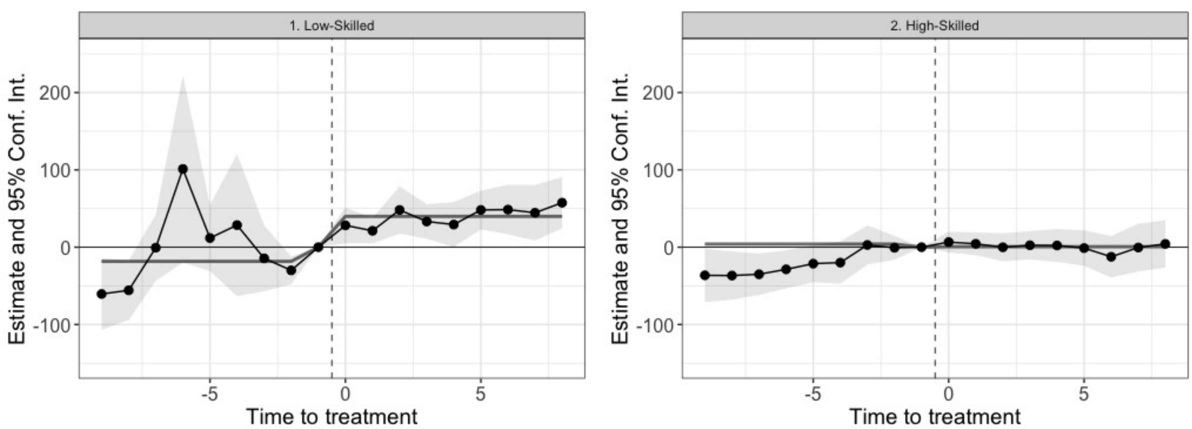
Again, to address the parallel trends assumption, we proceed by visually inspecting the pre-trend period. For y_{it}^{QUAL} Figure 3 displays no significant pre-trend irrespective of the level of skill. For the work speed measure y_{it}^{QUAN}, similar to the full sample, Figure 4 shows some periods further away from treatment violate the parallel trends assumptions, especially for high-skilled employees, which will be further addressed in Section 6. However, the findings for high-skilled agents on y_{it}^{QUAL} and for low-skilled agents y_{it}^{QUAN} do not appear to be driven by significant differences in the pre-period, since the pre-treatment average effect is practically zero in both specifications, as indicated by the grey line.

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



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). The grey line represents the average treatment effect in the pre and post-treatment periods.

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



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. The grey line represents the average treatment effect in the pre and post-treatment periods.

Besides the visual inspection of potential pre-trends, event-study graphs allow the inspection of dynamic treatment effects. Both the effects on y_{it}^{QUAL} of the high-skilled agents and on y_{it}^{QUAN} of the low-skilled agents show an immediate response in the first treatment month that remains relatively stable in the months following.

To further analyze skill differences, in Table A2, we split our sample based on the quartile of the ability scores (rather than the median) to highlight effect heterogeneity at the tails of the distribution. This can also speak for or against the question whether there are ceiling effects in performance. In that specification, we refer to agents as 'low', 'medium-low', 'medium-high' and 'high' skilled'. The findings - even though the sample restriction reduces statistical power - show that for service quality as an outcome, positive treatment effects are found only for those at the very top of the skill distribution, driving the initial results. For the high-skilled group, the positive effect is significantly stronger compared to the median split (0.080 versus 0.051). Negative effects on work speed, are driven by medium-low-skilled employees, while low, medium-high and high-skilled employees display no significant response to the bonus introduction for that outcome. This also provides evidence against ceiling effects in perfomance in the upper end of the performance distribution.

As explained in Section 3.1, the NPS (or y_{it}^{QUAL}) consists of three service quality components: Bad (score: 0-6), medium (score: 7-8) and good (score: 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 A3 we use the share of bad (Columns (1)-(3)), medium (Columns (4)-(6)) and good (Columns (7)-(9)) calls as dependent variable. The findings suggest that high-skilled agents reduced the number of bad calls while increasing the share of good calls. In contrast, low-skilled agents make more bad calls in response to the treatment.

Lastly, we can also examine how the resulting bonus levels differ by skill level. Because we do not observe bonus levels under the old system, we present descriptive results on bonus payments under the new system. Figure A5 in the Appendix shows that high-skilled individuals are more likely to get larger bonuses, whereas low-skilled individuals are more likely to get smaller bonuses. The figure also shows, however, that the tails of the distribution are far from zero. This is contrary to bonus distribution under the old scheme, where the lowest and highest bonus level are hardly awarded (see, Footnote 4).

5.3 Incentive pay 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 increased fatigue, reduced concentration and, consequently, a decline in the quality of calls. Furthermore, explicit performance incentives can serve as an alternative way to earn rents on their ability instead of working overtime. Following this argument, we would expect a more pronounced negative effect on overtime hours for high-skilled workers. While recent survey evidence on performance-related pay and working hours (Artz and Heywood, 2024; Green and Heywood, 2023) show rather an increase in working hours in response to such incentives, it is difficult to establish causality. In this section, we provide causal estimates for the effect of performance pay on working hours and explore possible heterogeneities between low- and high-skilled agents.

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 agents' (immediate) 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 4.79 (5.21) hours fewer than their contractual working hours. Regardless of other control variables, this gap shrank to 3.81 (4.26) hours per week in the nine months after the introduction.

Table 5 displays the estimated coefficients when regressing the average hours of overtime (per week) on the same set of control variables as in Table 3, as they may correlate with overtime and the probability of being subject to treatment. The results show, that on average, workers reduced weekly overtime by 1.51 hours in response to the individual bonus payment. Looking at the sub-samples (columns 2 and 3), this finding applies regardless of the level of skill.

Table 5: 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**	-1.51*	-1.66*
	(0.611)	(0.855)	(0.913)
Hours planned	-0.218***	-0.240***	-0.199***
	(0.036)	(0.059)	(0.043)
tenure	0.854	0.558	-1.24
	(0.530)	(0.827)	(0.928)
# Eval. Calls	0.039***	0.035***	0.045***
	(0.008)	(0.012)	(0.011)
Temp. Agent	0.202	0.937	-0.151
	(0.427)	(0.751)	(0.459)
Department dummies	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	2,396	1,178	1,218
R^2	0.546	0.574	0.569
Mean dependent Variable	-4.990	-5..213	-4.789
N Agents	303	152	151

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 A6 shows that the parallel trends assumption mostly holds for the main and sub-samples, especially in periods close to the treatment. 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 tend to avoid long hours to potentially avoid negative evaluations by customers, which have been documented for other performance outcomes (Collewet and Sauermann, 2017).[17]

[17] In line with that assumption, we find a negative correlation coefficient (-0.046) between overtime hours and NPS scores, which is consistent with our hypothesis. We find no meaningful association for average handling time.

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 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.[18] 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 A4, Figure A7 and Figure A8). With that, our baseline results amplify: The effect of the bonus introduction on y_{it}^{QUAL} increases from 0.051 to 0.062 for the high-skilled and the effect on y_{it}^{QUAN} from 37.7 to 54.0 for the low-skilled agents. Again, this implies that our initial estimates are a lower bound of the true treatment effect. Furthermore, when looking at the event-study designs, for y_{it}^{QUAN} (Figure A8), similar to the main results, the graphs imply a violation of the parallel trends assumption in some of the periods that are further away from the treatment. In the following section, we will therefore investigate potential violations of parallel trends further.

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

Violation of parallel trends Apart from visually inspecting potential violations of parallel trends in the framework of Sun and Abraham (2021), we use an alternative specification that allows for staggered treatment and treatment effect heterogeneity. Gardner (2022)'s approach provides a two-stage Difference-in-Difference estimator. 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. Similar to dynamic TWFE approaches like Sun and Abraham (2021), the estimator provides sensible results assuming parallel trends and no anticipation.

Significantly, the results displayed in Figure A9 show no significant violation of the parallel trends assumption in the six months leading up to treatment in any of the specifications. Violations that happen earlier, are unlikely to be related to the introduction of the treatment. Furthermore, the main post-treatment results increase in significance, as almost all of the coefficients are significant on the 5% level. Additionally, the findings displayed in Table A5 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. Compared to the baseline findings, the effect sizes for high-skilled in y_{it}^{QUAL} and low-skilled in y_{it}^{QUAN} are similar. However, given the alternative specification, the results additionally suggest significant negative treatment effects in terms of y_{it}^{QUAL} for low-skilled agents. Again, the event studies show no violation of parallel trends for that result.

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 individual performance pay, non-treated agents may adjust their efforts 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 A6 (Panel A) show that the inclusion of team-fixed effects does not qualitatively affect our main results. Again, the findings additionally imply significant negative treatment effects for low-skilled agents on y_{it}^{QUAL}.[19]

Additionally, the share of treated individuals per team may put peer pressure on untreated agents, as for instance, the communication about the newly introduced individual 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 A6 (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 individual 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 occur randomly. Furthermore, we consider the number of evaluated calls as the dependent variable, and the results in Table A7 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.[20] Compared to the main specification, the sample size increases by around 180 observations (or 5%). The results (Table A8) confirm a positive

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

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

treatment effect for high-skilled agents in terms of y_{it}^{QUAL} and negative treatment effects on y_{it}^{QUAN} for low-skilled employees. Similar to previous robustness tests, we also find significant negative treatment effects on the low-skilled in terms of service quality.

To further validate that our results are not driven by the choice of skill definition, instead of regression-based rankings (see, section 3.2), we also use a measure that is solely based on raw pre-treatment differences (instead of residuals) in service quality and work speed. The results displayed in Table A9 align with the baseline results. Furthermore, when relying on pre-treatment information only in the incentivized outcome y_{it}^{QUAL} to construct the skill measure, the findings for residualized (Table A10) and non-residualized (Table A11) definitions of skill align with the inital results. Contrary to all other skill definitions, the estimate for work speed among high-skilled agents becomes significant as well.

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 A12 - even though the findings for high-skilled agents and service quality are only significant on the 12% level -

[21] In our sample, we see a lot of turnover: Among the high-skilled (low-skilled) agents, 61.9% (67.0%) left the firm before the end of the observation period. Furthermore, 36.6% (35.1%) of high-skilled (low-skilled) agents left the firm in the six months leading up to the introduction. Following the treatment introduction, 25.3% (32.0%) of high-skilled (low-skilled) agents left the firm. In Figure A10, we also show the survival probabilities separately for low-skilled (dashed) and high-skilled (solid) agents. In line with the numbers reported above, low-skilled agents have slightly lower survival probabilities compared to high-skilled agents but which are not significantly different from each other (p-value=0.13). Overall, the findings are more consistent with slightly weaker sorting responses from low-skilled agents out of the company.

indicate qualitatively similar coefficients as in our baseline estimates and that 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 on average react to individual incentives set by management and increase their performance in terms of service quality to get a monetary bonus. This 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, or due to the incentive design. 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 individual bonus payment and display no effect on work speed. Low-skilled agents, on the other hand, did not improve service quality and even decreased work speed by 11%. Finally, contrary to recent empirical studies, the current study finds that under qualitative performance incentives, agents reduce overtime hours in response to individual 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. Our results imply that incentives can affect wage inequality within firms and potentially explain sorting effects of performance pay. 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 (2024): "Performance Pay and Work Hours: US Survey Evidence," *Oxford Economic Papers*, 76, 609–627.

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.

Baehre, S., M. O'Dwyer, L. O'Malley, and N. Lee (2022): "The use of Net Promoter Score (NPS) to predict sales growth: insights from an empirical investigation," *Journal of the Academy of Marketing Science*, 50, 67–84.

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 Work from Home," *Journal of Economic Perspectives*, 37, 23–50.

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.

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.

Ichino, A. and R. T. Riphahn (2005): "The Effect of Employment Protection on Worker Effort: Absenteeism during and after Probation," *Journal of the European Economic Association*, 3, 120–143.

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.

Mainar, I. G., C. P. Green, and M. N. Paniagua (2018): "The Effect of Permanent Employment on Absenteeism: Evidence from Labor Reform in Spain," *ILR Review*, 71, 525–549.

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

——— (2023): "Information, Incentives, and Attention: A Field Experiment on the Interaction of Management Controls," *The Accounting Review*, 98, 455–479.

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

Opitz, S., D. Sliwka, T. Vogelsang, and T. Zimmermann (2024): "The Algorithmic Assignment of Incentive Schemes," *Management Science*, forthcoming.

Otto, A. S., D. M. Szymanski, and R. Varadarajan (2020): "Customer satisfaction and firm performance: insights from over a quarter century of empirical research," *Journal of the Academy of Marketing Science*, 48, 543–564.

Roth, J., P. H. Sant'Anna, A. Bilinski, and J. Poe (2023): "What's Trending in Difference-in-differences? A Synthesis of the Recent Econometrics Literature," *Journal of Econometrics*, 235, 2218–2244.

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.

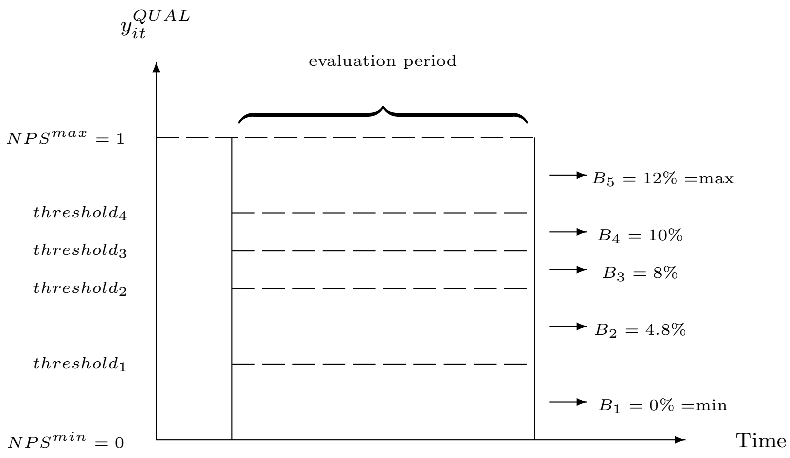
Unger, O., A. Szczesny, and M. Holderried (2020): "Does Performance Pay increase Productivity? Evidence from a Medical Typing Unit," *Management Accounting Research*, 47, 100649.

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

Online Appendix

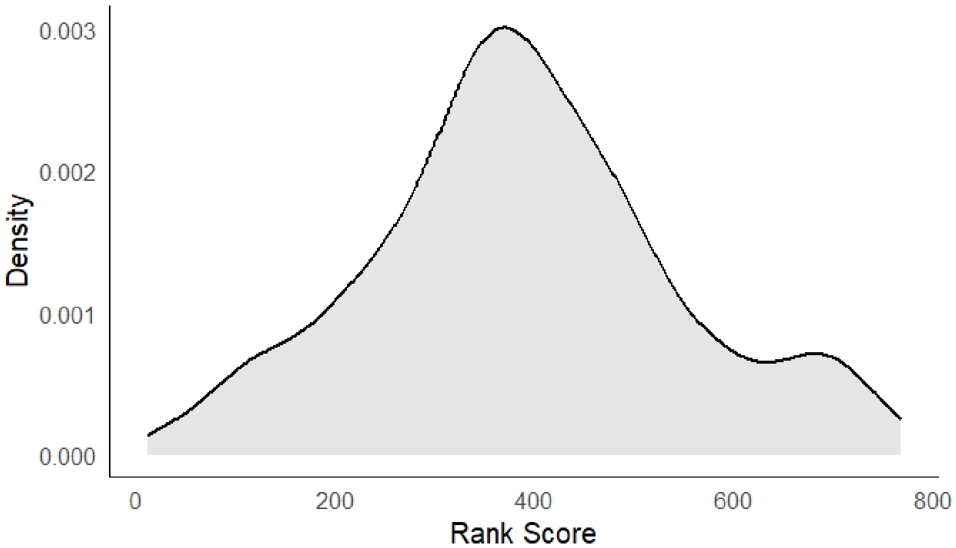
Clemens, Marco, and Jan Sauermann. 2024. "Making the Right Call: The Heterogeneous Effects of Individual Performance Pay on Productivity"

Figure A1: Design of the Bonus System



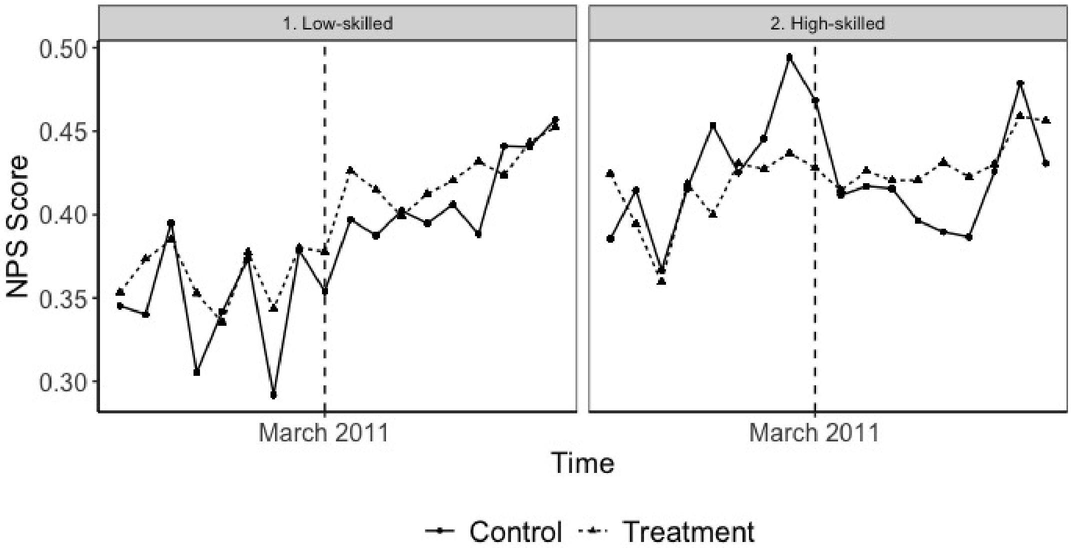
Notes: B_j denotes bonus levels as a percentage of an agent's gross wage in the bonus quarter when achieving a performance level of $threshold_j$.

Figure A2: Distribution of Total Rank Scores based on NPS and AHT



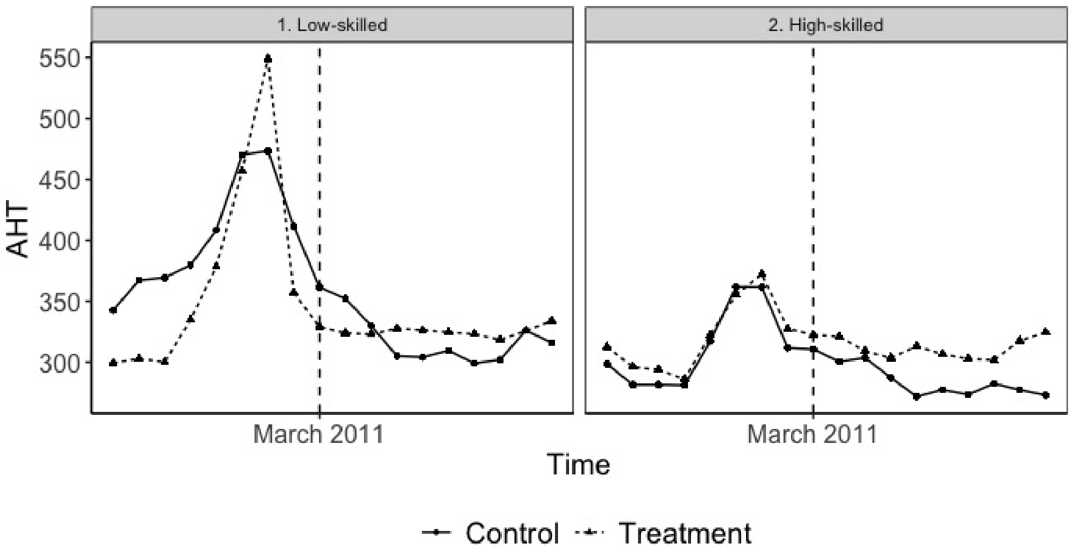
Notes: This figure displays the total rank scores based on residualized y_{it}^{QUAL} and y_{it}^{QUAN} in the six months leading up to the treatment, i.e. from October 2010 to March 2011. For a full description on the definition of skill, please re-visit section 3.2.

Figure A3: Evolution of Service Quality y_{it}^{QUAL} over Time



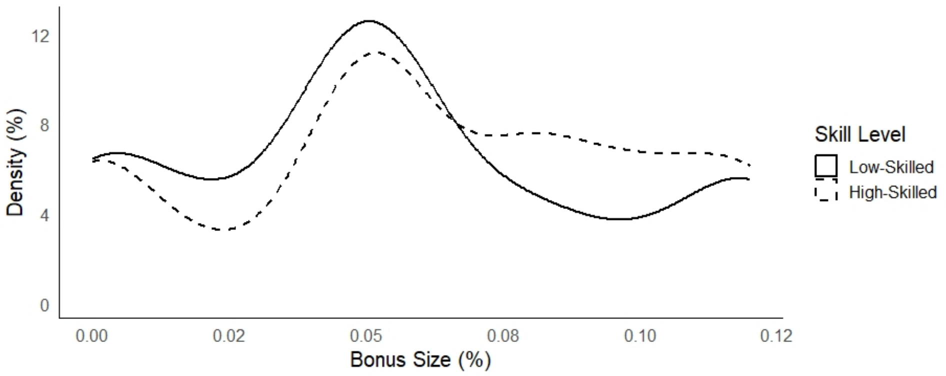
Notes: This figure displays $y_{i,t}^{QUAL}$ over time for treated and non-treated agents. The left-hand side shows the results for low-skilled and the right-hand sight for high-skilled agents. The last month before the treatment is indicated by the vertical dashed line. $y_{i,t}^{QUAL}$ is the service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

Figure A4: Evolution of Work Speed y_{it}^{QUAN} over Time



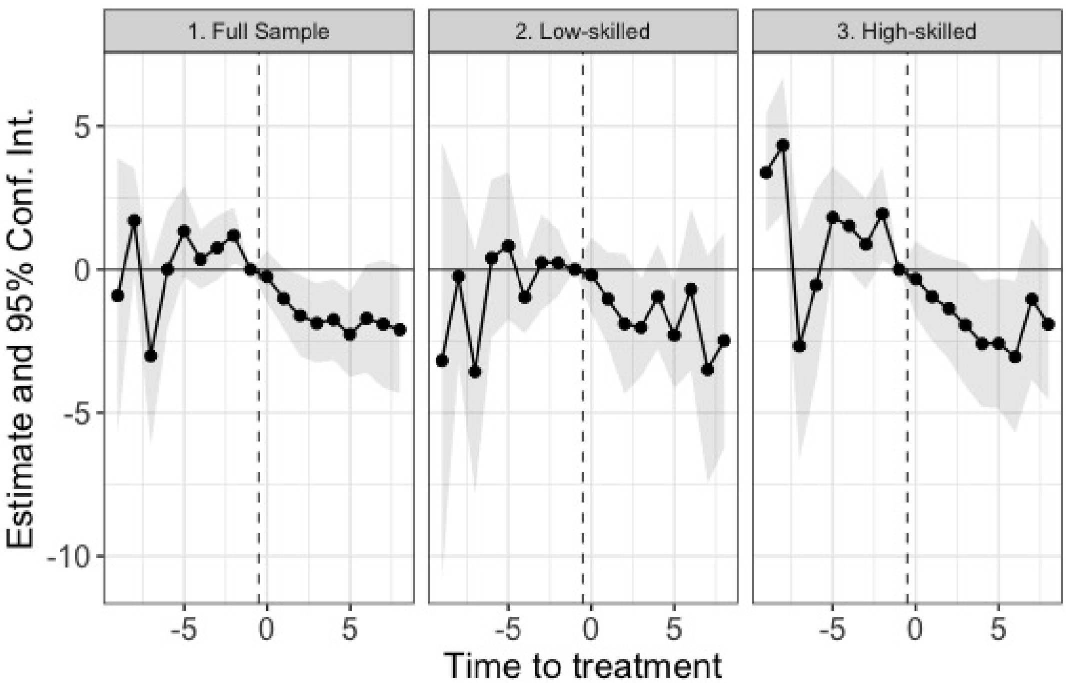
Notes: This figure displays y_{it}^{QUAN} for treated and non-treated agents over time. The left-hand side shows the results for low-skilled and the right-hand sight for high-skilled agents. The last month before the treatment is indicated by the vertical dashed line. $y_{i,t}^{QUAN}$ is the work speed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

Figure A5: Density by Skill Level



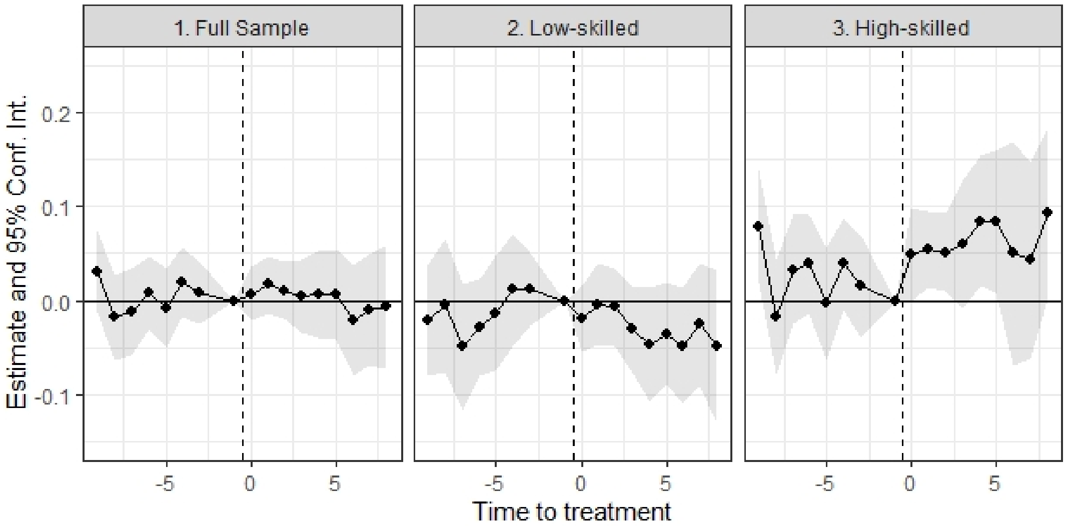
Notes: This figure displays the distribution of bonus pay based on the level of skill in the post-treatment period. Hence, the sample includes the periods between April 2011 and December 2011.

Figure A6: Event Studies of the Effect of the Bonus Introduction on Overtime



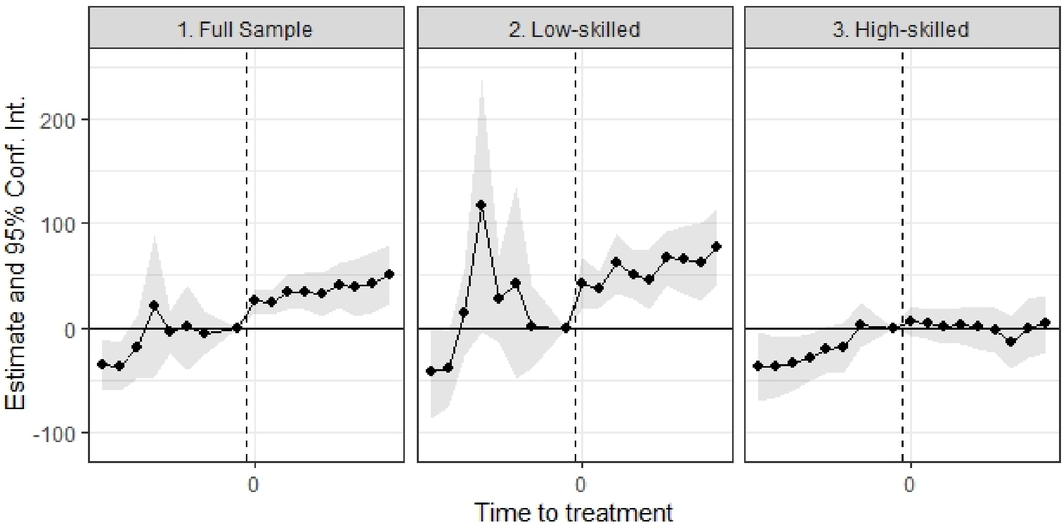
Notes: Event Study Analysis based on Sun and Abraham (2021) showing treatment effects on weekly overtime hours for each relative treatment period. The sample includes the time between October 2010 and December 2011. Overtime hours are defined as the difference between actual and contractual working hours.

Figure A7: The Effect of the Bonuses on Service Quality (y_{it}^{QUAL}) - Event Studies with Adjusted Reference Period



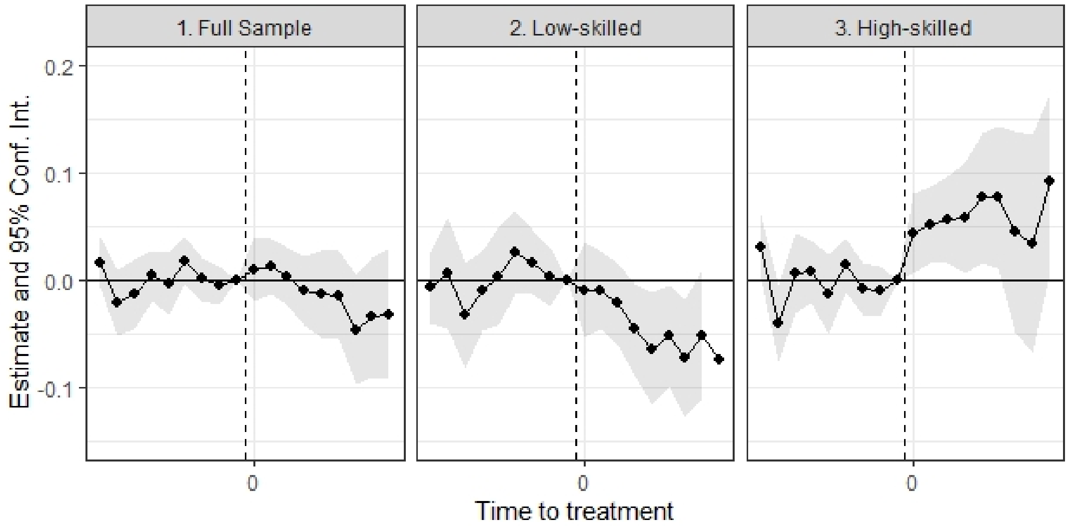
Notes: Event Study Analysis based on Sun and Abraham (2021) showing treatment effects on $y_{i,t}^{QUAL}$ for each relative treatment period. The sample includes the time between October 2010 and December 2011. The estimates use *t-1* and *t-2* as reference category. $y_{i,t}^{QUAL}$ is the service quality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

Figure A8: The Effect of the Bonuses on Work Speed (y_{it}^{QUAN}) - Event Studies with Adjusted Reference Period) - Event Study

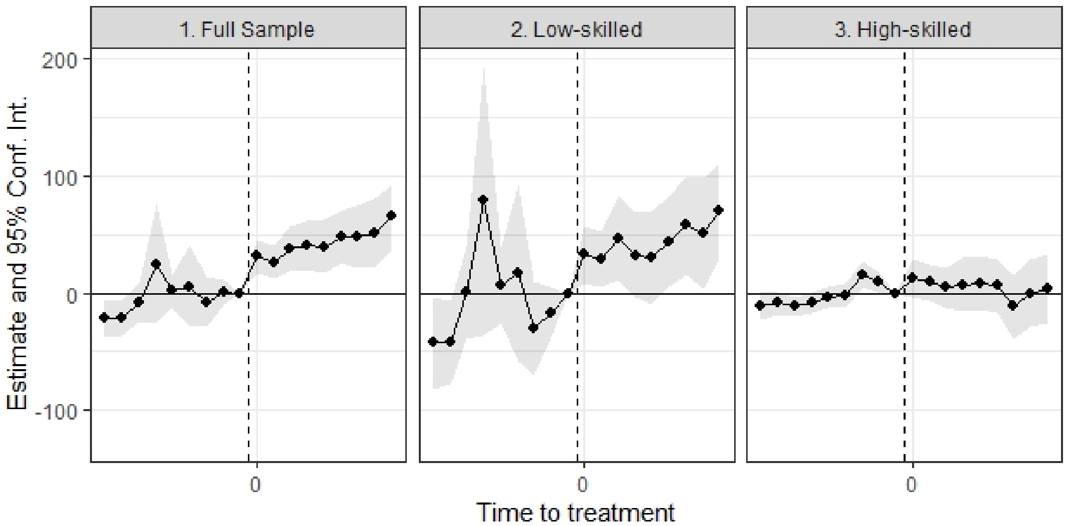


Notes: Event Study Analysis based on Sun and Abraham (2021) showing treatment effects on $y_{i,t}^{QUAN}$ for each relative treatment period. The sample includes the time between October 2010 and December 2011. The estimates use *t-1* and *t-2* as reference category. $y_{i,t}^{QUAN}$ is the work speed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

Figure A9: Two-Stage DiD Estimates based on Gardner (2021)



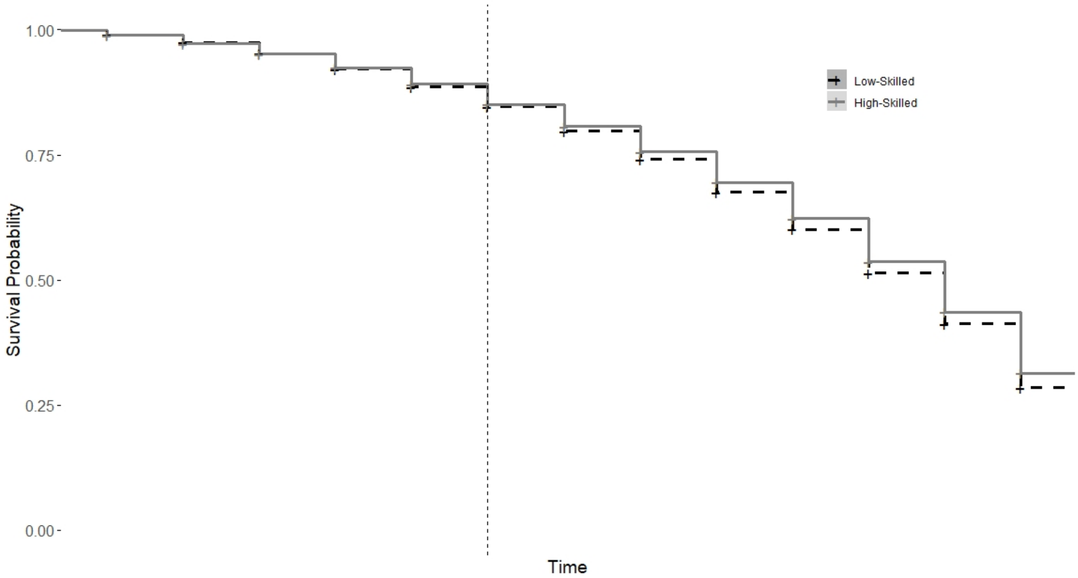
(a) Service Quality - y_{it}^{QUAL}



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

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 A10: Survival Probabilities by Skill-Level



Notes: This figure displays survival probabilities for high-skilled (solid line) and low-skilled (dashed line) agents over time. The last month before the treatment is indicated by the vertical dashed line. The sample includes the time between October 2010 and December 2011.

Table A1: The Effect of the Bonus Introduction on Service Quality and Work Speed - Restricted Sample

Dep. Variables	Service Quality (y_{it}^{QUAL})				Work Speed (y_{it}^{QUAN})			
	TWFE		Sun and Abraham (2021)		TWFE		Sun and Abraham (2021)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Performance Pay	-0.0008	-0.002	0.010	0.001	46.5***	47.7***	22.3***	23.6**
	(0.013)	(0.013)	(0.019)	(0.020)	(9.04)	(9.06)	(8.17)	(9.12)
Hours Planned		0.0005		0.0002		0.434		0.258
		(0.0009)		(0.0009)		(0.510)		(0.525)
Tenure		0.007		0.010		0.428		1.92
		(0.010)		(0.010)		(6.74)		(6.84)
# Eval. Calls		0.000		0.0001		-0.631***		-0.644***
		(0.0002)		(0.0002)		(0.151)		(0.163)
Temp. Agent		-0.0005		0.0004		16.3		18.0*
		(0.014)		(0.016)		(10.6)		(10.5)
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,879	2,879	2,879	2,879	2,879	2,879	2,879	2,879
R^2	0.265	0.285	0.285	0.304	0.593	0.635	0.611	0.650
Within R^2	0.016	0.042	0.027	0.053	0.137	0.226	0.058	0.153
Mean dep. variable	0.404				330.77			
N Agents	346							

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 restricted samples excludes agents that became subject to the bonus and moved out of temporary employment at the same time. 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: 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-Low	Medium-High	High	Low	Medium-Low	Medium-High	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Performance Pay	-0.038	0.023	0.036	0.080**	-30.9	30.8**	10.5	-1.35
	(0.048)	(0.036)	(0.039)	(0.035)	(23.8)	(14.3)	(10.5)	(14.1)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	702	979	1,034	802	702	979	1,034	802
R^2	0.413	0.389	0.336	0.328	0.798	0.612	0.748	0.780
Mean Dependent variable	0.365	0.390	0.410	0.440	381.711	347.574	328.788	283.552
N Agents	97	97	97	97	97	97	97	97

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 (5), medium-skilled agents in Columns (2), (3), (6) and (7), and high-skilled agents in Columns (4) and (8). 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: 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	0.048*	-0.052*	-0.016	-0.023	0.0008	0.0003	-0.025	0.051*
	(0.022)	(0.029)	(0.030)	(0.019)	(0.028)	(0.026)	(0.019)	(0.025)	(0.027)
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	1,681	1,836	3,517	1,681	1,836	3,517	1,681	1,836
R^2	0.325	0.373	0.323	0.270	0.297	0.285	0.241	0.272	0.269
Mean Dep. Variable	6.177	6.479	5.870	6.142	5.829	6.461	3.134	2.735	3.539
N Agents	388	194	194	388	194	194	388	194	194

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 A4: Baseline Estimates With Adjusted Reference Category

Dependent Variables:	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	0.004	-0.026	0.062***	34.3***	54.0***	0.868
	(0.015)	(0.018)	(0.023)	(8.26)	(10.6)	(8.04)
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	1,681	1,836	3,517	1,681	1,836
R^2	0.299	0.346	0.306	0.598	0.647	0.751
Mean Dependent variable	0.405	0.363	0.448	333.24	349.576	316.620
N Agents	388	194	194	388	194	194

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 A5: Estimates Based on Gardner (2022)

Dependent Variables:	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 (Gardner, 2022)	-0.009	-0.039***	0.059***	40.6***	41.0***	5.75
	(0.011)	(0.015)	(0.017)	(8.78)	(14.9)	(8.93)
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	1,681	1,836	3,517	1,681	1,836
R^2	0.0004	0.010	0.031	0.034	0.019	0.002
Mean Dependent variable	0.405	0.363	0.448	333.24	349.576	316.620
N Agents	388	194	194	388	194	194

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 - 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	-0.042*	0.053**	17.3**	37.2***	0.874
	(0.019)	(0.025)	(0.026)	(7.41)	(11.0)	(9.65)
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	1,684	1,814	3,498	1,684	1,814
R^2	0.316	0.380	0.326	0.625	0.690	0.779
Mean Dependent Variable	0.405	0.362	0.449	344.9	333.240	316.297
N Agents	379	190	189	379	190	189
Panel B						
Performance Pay	-0.002	-0.028	0.055**	19.3***	33.5***	-2.33
	(0.018)	(0.024)	(0.026)	(6.78)	(9.25)	(8.96)
$\overline{T_{post}}$	0.020	-0.003	0.010	-31.5**	-36.1	-40.9*
	(0.026)	(0.037)	(0.039)	(14.9)	(24.7)	(21.2)
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	1,446	1,474	2,920	1,446	1,474
R^2	0.324	0.377	0.327	0.636	0.631	0.773
Mean Dependent Variable	0.407	0.365	0.449	332.319	348.201	316.297
N Agents	355	177	178	355	177	178

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 A7: The Effect of the Bonus Introduction on Evaluated Calls

Dep. Var.	Evaluated calls		
	Full sample	Low-skilled	High-skilled
	(1)	(2)	(3)
Performance Pay	4.05*	3.55	5.22
	(2.30)	(3.48)	(3.18)
Control variables	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	3,517	1,681	1,836
R^2	0.754	0.760	0.774
Mean Dependent Variable	15.453	15.043	15.870
N Agents	388	194	194

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 A8: Skill Differences (1) - Broader Skill Measure

	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	-0.042*	0.046*	25.5***	33.7***	5.21
	(0.018)	(0.023)	(0.025)	(7.64)	(10.9)	(8.75)
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	1,693	2,004	3,697	1,693	2,004
R^2	0.331	0.362	0.324	0.615	0.644	0.713
Mean Dependent Variable	0.405	0.4505	0.50	333.240	357.096	312.462
N Agents	484	242	242	484	242	242

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 A9: Skill Differences (2) - Raw Rankings

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	-0.008 (0.018)	-0.038* (0.020)	0.062** (0.031)	25.5*** (7.70)	23.1** (11.5)	7.26 (8.68)
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	1,699	1,818	3,517	1,699	1,818
R^2	0.299	0.345	0.298	0.598	0.586	0.758
Mean dependent Variable	0.405	0.353	0.461	333.240	365.339	298.849
N Agents	388	195	193	388	195	193

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. Skill differences are defined based on pre-treatment rankings of both $y_{i,t}^{QUAL}$ and $y_{i,t}^{QUAN}$. 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 A10: Skill Differences (3) - Measure Based on NPS (Residualized Rankings)

	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	-0.008 (0.018)	-0.040* (0.023)	0.057** (0.027)	25.5*** (7.70)	30.5*** (9.81)	16.5 (11.1)
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	0.370	0.288	0.598	0.601	0.694
Mean Dependent Variable	0.405	0.333	0.479	333.24	326.53	340.18
N Agents	388	194	194	388	194	194

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. Skill is defined based on the residualised service quality in the six months prior to the treatment introduction. $y_{i,t}^{QUAL}$ 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 A11: Skill Differences (4) - Skill Measure Based on NPS (Raw Rankings)

	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	-0.008	-0.048*	0.058**	25.5***	23.8**	23.5**
	(0.018)	(0.025)	(0.026)	(7.70)	(9.19)	(11.6)
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	3,517	1,702	1,815	3,517	1,702	1,815
R^2	0.299	0.347	0.266	0.598	0.638	0.609
Mean Dependent Variable	0.405	0.318	0.486	333.24	326.30	339.72
N Agents	388	194	194	388	194	194

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. Skill is defined based on the service quality in the six months prior to the treatment introduction. $y_{i,t}^{QUAL}$ 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 A12: 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	-0.008	-0.029	0.041	22.5***	36.8***	-4.16
	(0.019)	(0.024)	(0.026)	(7.79)	(9.60)	(9.22)
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	1,424	1,524	2,948	1,424	1,524
R^2	0.299	0.350	0.324	0.542	0.620	0.717
Mean Dependent Variable	0.404	0.369	0.441	334.189	350.209	316.770
N Agents	249	126	123	249	126	123

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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.