

Human capital, incentives, and performance outcomes

Citation for published version (APA):

Sauermann, J. (2013). *Human capital, incentives, and performance outcomes*. ROA.

Document status and date:

Published: 01/01/2013

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Human Capital, Incentives, and Performance Outcomes

© Jan Sauermann, Maastricht 2012

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission in writing, from the author.

Published by ROA
Postbus 616
6200 MD Maastricht

ISBN: 978-90-5321-504-3

Printed in the Netherlands by Océ Business Services

Human Capital, Incentives, and Performance Outcomes

Proefschrift

ter verkrijging van de graad van doctor aan
de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof. dr. L.L.G. Soete
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
vrijdag 25 januari 2013 om 10.00 uur

door

Jan Sauermann

Promotor

Prof. dr. A. de Grip

Co-promotor

dr. I. Sieben (Tilburg University)

Beoordelingscommissie

Prof dr. T. Dohmen (voorzitter)

Prof dr. W. Hassink (Utrecht University)

Prof dr. B. ter Weel

Acknowledgements

I am very grateful to Andries De Grip for giving me the possibility to pursue my doctoral studies at Maastricht University. Throughout the years, Andries offered continuous advice and always listened to my questions, whether it was about research, giving presentations, or on life in the Netherlands in general. For invaluable feedback and critical assessment of my papers and the final manuscript of this thesis, I would like to thank Inge Sieben. The Research Centre for Education and the Labour Market (ROA) at Maastricht University provided an excellent place for doing research. I benefitted substantially from feedback on drafts of my research papers, and enjoyed the great atmosphere at ROA.

Furthermore, I would like to thank seminar and conference participants who provided feedback on earlier drafts of my work as well as the assessment committee of this thesis. This thesis would not have been possible without the help of Patrick Smits and Stefan Terpstra, for whose cooperation and trust concerning data privacy I would like to thank; and Baukje Janssen, Dominique Gilissen, Ivo Smits, and Daniel Harst who delivered the data and always helped me with the data.

Lastly, I would like to thank all my family, friends from Maastricht and elsewhere for their continuous support before, during, and after my doctoral studies.

Stockholm
December, 2012

Jan Sauermann

Contents

1	Introduction	1
1.1	Motivation	2
1.2	Aim of this thesis	4
1.3	Outline	5
2	Data	9
2.1	The use of personnel data for economic research	10
2.2	Data used in this thesis	11
2.2.1	Organisation of the firm	11
2.2.2	Performance measures	12
2.3	Comparison to other firms and other sectors	14
3	The role of peers in estimating tenure-performance profiles	15
3.1	Introduction	16
3.2	The firm	19
3.2.1	Information on the workplace	19
3.2.2	Sample selection and descriptive statistics	20
3.3	Estimating tenure-performance profiles	25
3.3.1	Baseline results	25
3.3.2	Investments in learning on-the-job	27
3.3.3	Team composition and tenure-performance profiles	28
3.4	Robustness analyses	31
3.4.1	Turnover and selection of workers	31
3.4.2	Assignment to teams and shifts	33
3.4.3	Substitution between quality and quantity of calls	34
3.4.4	Working hours and the shape of tenure-performance profiles	37
3.5	Conclusions	39
4	The effects of training on own and co-worker productivity	41
4.1	Introduction	42
4.2	Context of the experiment	45
4.2.1	Organisation of the department and worker tasks	45
4.2.2	Training purpose, contents, and organisation	47

Contents

4.2.3	The field experiment	48
4.2.4	Measuring productivity	52
4.3	Empirical analysis	53
4.3.1	Causal effect of training on individual performance	53
4.3.2	Baseline results	54
4.3.3	The role of social interaction	58
4.3.4	Returns to training	60
4.4	Additional evidence	62
4.4.1	First-week effects	62
4.4.2	Personnel turnover and training effects	63
4.4.3	Effects on call quality	64
4.4.4	Heterogeneous treatment effects	66
4.5	Conclusions	67
5	Under pressure: Supervisor’s role in implicit targets and worker effort	69
5.1	Introduction	70
5.2	Theoretical framework	71
5.3	The firm	75
5.4	Empirical approach	76
5.5	Results	78
5.5.1	Descriptive evidence	78
5.5.2	Estimation results	80
5.6	Conclusion	89
6	The heterogeneous effects of bonus pay on performance outcomes	91
6.1	Introduction	92
6.2	The firm and its incentive schemes	94
6.2.1	The firm	94
6.2.2	Explicit incentives in the firm	94
6.2.3	Performance outcomes	97
6.3	The effects of incentives on performance	99
6.3.1	Estimation sample	99
6.3.2	Descriptive results	102
6.3.3	Estimation strategy	104
6.3.4	Estimation results	105
6.4	Robustness analyses	114
6.4.1	Seasonality and confounding factors	114
6.4.2	Sorting and turnover	118
6.5	Conclusion	120
7	Conclusion	121
7.1	Summary	122
7.2	Contributions and implications for human resource management	124
	Bibliography	127

Summary in Dutch	135
Biography	139
ROA Dissertation Series	141

List of Figures

1.1	Overview of topics in this thesis	4
3.1	Kernel density plot for tenure	21
3.2	Non-parametric and logarithmic tenure-performance profiles	22
3.3	Kernel density plot for performance	24
3.4	Non-parametric tenure-quality profile	35
4.1	Overview of the field experiment	49
4.2	Average performance of the treatment group before and after the training	54
4.3	Development of the treatment effect on performance over time	57
4.4	Estimated payoff (in euros)	62
5.1	Percentile differences in performance y_{jt} and residual performance $\hat{\epsilon}_{it}$	79
5.2	Estimated effect of agent-team and team rankings by decile	84
6.1	Design of the bonus system	96
6.2	Distribution of bonuses paid in the first performance pay quarter	101
6.3	Service quality before and after the introduction of performance bonuses	103
6.4	Average handling times and shares of problem solved before and after the introduction of performance bonuses	108
6.5	Distribution of service quality by payment scheme	111
6.6	Distribution of average handling times and shares of problem solved by payment scheme	113
6.7	The effect of shorter sample periods on the estimated treatment effect	117
6.8	The effect of placebo treatments on the estimated treatment effect	118
6.9	Turnover rates over time	119

List of Tables

3.1	Descriptive statistics (main variables)	23
3.2	The effect of tenure on performance	26
3.3	Descriptive statistics (other variables)	29
3.4	The effect of peers' experience on individual performance	31
3.5	The effects of tenure on new hires	32
3.6	Selection into teams	34
3.7	The effect of tenure on quality	36
3.8	Determinants of performance using different tenure definitions	38
4.1	Selection into the experiment	46
4.2	Descriptive statistics of teams' first and second training groups (treatment group only)	51
4.3	Average performance post-treatment period	55
4.4	Direct treatment effect	56
4.5	Estimation of externalities	60
4.6	Pre-/post-differences in performance for other training programmes	63
4.7	Estimated treatment effect and estimation samples	64
4.8	The effect of training participation on call quality	65
4.9	Direct treatment effect with interaction terms	67
5.1	Descriptive statistics	75
5.2	OLS results of previous individual ranking	81
5.3	OLS results of previous team ranking	82
5.4	Joint estimation of individual and team ranking	83
5.5	Spline estimation of the relation between ranking and effort – 1	85
5.6	Spline estimation of the relation between ranking and effort – 2	87
5.7	Spline estimation by past performance quintiles	88
5.8	One-step estimation	89
5.9	First-stage regression	90
6.1	Descriptive statistics	100
6.2	Unconditional differences before and after the introduction of performance pay	102
6.3	The effect of performance pay on service quality	106

List of Tables

6.4	The effect of the introduction of performance bonuses on average handling time and share of problems solved	109
6.5	Heterogenous effects of the introduction of performance bonuses on performance outcomes	112
6.6	Seasonal effects and performance pay	114
6.7	Sample length and the estimated treatment effect	115
6.8	The effect of performance pay on service quality and individual time trends	119

1 Introduction

1.1 Motivation

What drives workers' performance? What can firms do and which instruments can they apply to improve the performance of their employees? These questions are central to personnel economics. In a simple theoretical framework, a worker's performance is determined by the effort the worker puts into the task. The workers, in turn, choose an optimal level of effort which depends on their ability as well as on the incentives set by the firm's management (see e.g. Lazear, 2000). This basic framework provides the rationale for the analyses in this thesis. It also shows that performance provided by workers is not necessarily stable over time. In general, individuals adjust their provision of effort to the settings provided by the firm. Firms can choose from a variety of human resource (HR) practices in order to stimulate the performance of their employees (see Lazear, 1995; Ichniowski *et al.*, 1997). The list of potential HR practices is long, and depends on both the type of job and the type of firm. Examples of HR practices that have been studied in the economics literature are turnover-related instruments (e.g. recruitment; selection; and turnover policies), work organisation-related policies (e.g. teams and team composition; flexible work arrangements; hierarchies), remuneration (e.g. performance pay, non-monetary incentives), human capital development (e.g. training; skills improvement of workers), and labour-relations related instruments (e.g. unionisation; works councils; employment protection).

In this thesis, I analyse two ways to stimulate workers' effort and performance: human capital development and incentives. These two instruments leverage worker performance in different ways. Learning is assumed to increase workers' human capital and thereby workers' performance, whereas incentives attempt to increase the optimal effort level chosen by workers and thereby affecting their performance. Human capital development focuses on improving workers' skills through formal or informal learning. Bassanini *et al.* (2007) report that close to 40% of the employees in the Netherlands participate at least once in training, mostly paid for by the employer. Borghans *et al.* (2007) show that the largest part of the investments in human capital refers to informal learning. The effect of learning on economic outcomes such as wages and performance is based on human capital theory (Mincer, 1958; Becker, 1962). It is argued that firms invest in human capital, for example by providing training courses, if the expected return exceeds the costs of the investment. This basic theory can be applied not only to formal training but also informal learning, or learning by doing. A vast literature examines the returns to work-related training (e.g., Bassanini *et al.*, 2007), as well as the returns to informal learning (e.g., Shaw and Lazear,

2008). This literature finds that there are large returns to both participation in work-related training (*formal training*) and job tenure (*informal learning*). Estimation of the returns to learning, however, is cumbersome because of a severe selection bias, which can arise because certain types of workers are more likely to participate in training programmes or have longer tenures. In addition, this literature has paid less attention to incorporating the effects of peers on the returns to learning. In the last decade, a growing literature has emphasised the role of peers in providing higher worker performance (e.g., Sacerdote, 2001; Zimmerman, 2003; Falk and Ichino, 2006; Mas and Moretti, 2009, Bandiera *et al.*, 2010). Peers can affect their co-workers' performance in two ways: by applying pressure to perform better (*peer pressure*) or by providing knowledge (*spillover effects*). Analogous to the estimation of returns to training, however, peer effects are difficult to estimate because of endogeneity (Manski, 1993).

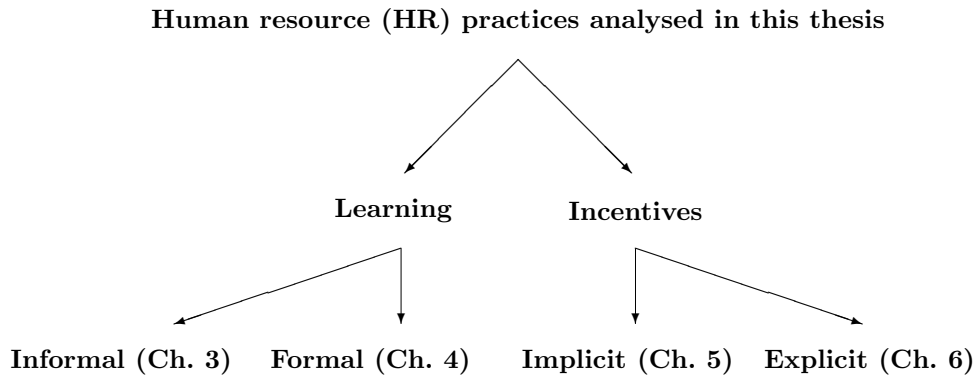
A second important HR practice involves incentives, which are used by firms to elicit optimal effort levels. Similar to formal and informal learning in firms, this instrument is widely used: Gielen *et al.* (2010) report that in 2001 39% of Dutch firms used performance pay as an HR instrument; 41% of call agents' pay in Dutch call centres is subject to incentives (Holman *et al.*, 2007).

The literature on the effects of providing monetary incentives that are paid upon achieving predefined performance thresholds shows that performance bonuses have substantial positive effects on workers' effort provision (e.g., Lazear, 2000; Shearer, 2004; Shi, 2010; Heywood *et al.*, 2011). This suggests that providing performance-related bonuses to workers can increase their effort and thus increase overall firm productivity. However, incentives can also be implicit, which is usually less well defined than explicit incentives, for example, promotions or dismissals if workers overperform or underperform, respectively. The literature on implicit incentives argues that workers can be incentivised by the firm's dismissal decisions (Lazear and Rosen, 1981) or promotion decisions, which might be driven by workers' underperformance or overperformance, respectively. One way in which firms can achieve this is through managers who monitor and evaluate their subordinates and thereby affect worker performance. This has recently been analysed by Lazear *et al.* (2012), who estimate the overall productivity effect of managers, and Barankay (2011a,b), who shows that providing individual ranking information affects workers' effort provision.

1.2 Aim of this thesis

This thesis aims at contributing to the literature on the returns to learning as well as the literature on incentives. In particular, Chapter 3 and 4 contribute to the literature on the effect of informal and formal learning on workers' performance, whereas Chapter 5 and 6 contribute to the literature on the effect of implicit and explicit incentives on workers' performance. Figure 1.1 gives an overview of the topics covered in Chapters 3 to 6.

Figure 1.1: Overview of topics in this thesis



One key feature of the studies of this thesis are the data. The data used in this thesis allow to follow individual workers of a particular firm over time. All four studies build on unique performance data of a call centre that contain detailed information of different measures of performance at the level of the individual worker. Having longitudinal information on individual performance as well as information on agents' tenure and formal training programmes, the data allow to assess the extent of formal and informal learning. More in general, these panel data allow to estimate the effects of events, such as the effect of participation in a training programme, and the effect of worker and team rankings on the individual workers' performance.

Chapters 3 and 4, which focus on informal and formal learning, contribute to the human capital literature by using personnel data that enable to assess the role of informal learning and formal training in the firm, respectively. The nature of the data does not only allow to quantify the effects of these types of learning, but also to compare the relative size of the two types of investments for newly hired workers (Chapter 3). A key element of the two studies is that I explicitly analyse the effect

that a worker's peers have on the human capital development of workers. For this purpose, Chapters 3 and 4 exploit quasi-exogenous and exogenous variation of peer composition. This has important implications for firms, as firms might exploit these externalities by, for instance, strategically assigning inexperienced workers to teams with more experienced peers.

The studies in Chapters 5 and 6 contribute to the literature of incentives in firms. The data used in this thesis allow to estimate the effects that rankings have on the workers' choice of effort and thus on their performance. Furthermore, Chapter 6, which analyses the effects of performance pay on workers' effort, shows the importance of the design of incentive schemes. This study shows that incentive pay can have negative effects for high-performing workers when the incentive scheme is not appropriately designed to account for differences in performance.

The primary purpose of the studies in this thesis is not to assess the relative importance of different HR practices. Rather, the studies provide an analysis of practices related to human capital development and incentives that are widely used in firms.

1.3 Outline

The remaining chapters of this thesis are structured as follows. Chapter 2 provides an overview of the data used throughout this thesis. Chapters 3 and 4 deal with informal and formal learning in the workplace, respectively, and the question of how externalities can affect these two types of learning. Chapter 3 analyses the importance of informal learning for performance outcomes and whether this learning is affected by workers' peers. Chapter 4 analyses whether formal training courses affect worker performance. Again, special emphasis is placed on the question whether there are peer effects in learning. Chapters 5 and 6 analyse how incentives affect worker performance. Chapter 5 discusses whether implicit targets that can impact promotion or dismissal decisions affect performance. Chapter 6 analyses whether explicit targets that are related to bonus payments result in higher performance and potentially distorting effects. Chapter 7 concludes the thesis and briefly discusses its main contribution to research in its field.

Chapter 2. Data

Chapter 2 gives an overview of the data used throughout this thesis. All chapters of this thesis build upon data collected at the call centre of a multi-national telecom-

munications company in the Netherlands. These data contain rich information on individual worker performance for any of the workers' working weeks. The data thus comprise a large panel of individual-level performance information.

Besides rich information on performance outcomes, the data allow for the identification of the various effects that are tested in this thesis. Chapter 5 uses longitudinal information to assess the effect of past rankings on current performance outcomes. Chapter 3 exploits quasi-exogenous peer assignment to identify peer effects in informal learning. Chapter 6 exploits the introduction of performance bonuses that depend on predefined performance targets to determine the effects of performance pay on performance outcomes. Chapter 4, which estimates the effects of participation in a training programme on performance, employs a field experiment with random assignments to treatment and control groups.

Chapter 3. Tenure–performance profiles: informal learning and peer effects

In Chapter 3, I analyse the tenure–performance profiles of newly hired call agents and the extent to which these profiles are affected by the agents' peers. The analysis of tenure–performance profiles originates from the literature on estimating the effects of tenure on wages (Lazear, 1979; Baker *et al.*, 1994; Altonji and Williams, 2005). One interpretation of upward sloping tenure–wage profiles is that they reflect an increasing stock of human capital, which is rewarded by higher wages (Mincer, 1958). An alternative way of estimating the relation between tenure/experience and performance is to use data from personnel records. These data are less prone to worker self-selection and often have direct measures of output (Shaw and Lazear, 2008). In Chapter 3, I use this approach to estimating tenure–performance profiles.

The results show that call agents have a very steep tenure–performance profile in the beginning of their job at the call centre. However, the slope flattens considerably after about six to eight months. The tenure–performance profile can be very well described by a logarithmic function. An increase in tenure of 10% leads to an increase of 5.5% of one standard deviation in performance. Agents' increase in performance in the first year of working in this job translates to a learning investment by the firm of about 162 hours. This suggests that, in terms of investment, informal learning is more important than formal training, since over the same period the firm's investments in formal training are much lower and accumulate to an average of only 84 hours.

Chapter 3 extends the literature on tenure–performance profiles by allowing for peer effects in learning. Barron *et al.* (1989) show that learning investments are not

merely the result of accumulating one's own experience, but also involve watching other workers performing their tasks or by receiving information from peers on how to perform well. This chapter shows that workers significantly benefit from working with more experienced peers. Estimation results show that an increase in average tenure of a worker's team mates of one standard deviation relates to an increase in worker performance of 14% of one standard deviation. This shows that it can be beneficial for firms to place new workers in teams with more experienced agents, thereby facilitating knowledge spillovers between workers, even in situations where workers are not involved in joint team work.

Chapter 4. The effects of training and externalities from training

Chapter 4 analyses the effect of training participation on worker performance by means of a field experiment. In this field experiment, agents were randomly assigned to training groups, allowing for the identification of causal effects of training on worker performance.

Unbiased estimation of the impact of work-related training on productivity is important to assess the role of training in the development of human capital (e.g., Bassanini *et al.*, 2007). Previous studies find mixed results, mostly due to the method applied, as well as the measure of productivity used. A common result is that the estimated effect of training on wages or performance decreases substantially once selection into treatment is appropriately controlled for (Bartel, 2000; Black and Lynch, 2001; Bassanini *et al.*, 2007; Leuven and Oosterbeek, 2008). This study uses a field experiment where the causal effects of treatment on outcome variables can be established. I show that workers' performance increases by 10% after participating in the training programme ('direct' effect). Furthermore, I show that workers who did not participate in the training but work with trained peers also benefit from the training ('indirect' effect). I find that a 10 percentage point increase in the share of treated agents leads to a performance increase of 0.51% for their untrained teammates.

Chapter 5. Implicit incentives

Chapter 5 analyses the role team leaders play in the performance of individual workers. In most organisations, workers are organised in teams. One purpose of teams of individually working employees is that a team leader, who is present in the workplace, can efficiently monitor and evaluate the workers in his or her team. This chapter

1 Introduction

focuses on the question whether supervisor pressure can be applied to subordinates to provide optimal effort, and thus to ensure that implicit incentives are functioning.

While a substantial literature has analysed the role of explicit incentives in eliciting optimal effort levels, there is only limited knowledge about the role of supervisors. An exception is the paper of Lazear *et al.* (2012), which examines and quantifies the role of managers on the work floor. Furthermore, Blanes i Vidal and Nossol (2011) and Barankay (2011a,b) show how the revelation of rankings within a piece rate pay distribution affects worker effort.

I estimate the effect of supervisor pressure, which is proxied by the agent's as well as the team's ranking, on worker performance. I find that the effect of individual rankings of past performance on current performance is much stronger than the effect of team rankings. The effect, however, is strongest among agents with the highest rankings in the previous week, while it is about half as strong for agents at lower ranks.

Chapter 6. Explicit incentives: Bonus pay and performance

Chapter 6 analyses the performance effects of bonus pay. To estimate these, I exploit the introduction of performance bonuses that depend on agent performance and predefined performance targets. The call centre management introduced individual performance bonuses to improve customer satisfaction with the agents' services. Under this bonus scheme, agents receive a monetary bonus of up to 12% upon achieving predefined performance thresholds.

In accordance with the literature, I find that the workers, on average, react to the newly introduced performance pay by increasing their performance. This result is in line with findings from laboratory and field studies for other industries (Lazear, 2000; Shearer, 2004; Shi, 2010; Heywood *et al.*, 2011). Estimation results show that workers who are lower in the performance distribution before the introduction of the performance pay scheme drive the performance improvement after the bonus is introduced, whereas the bonus does not affect the behaviour of workers who would have outperformed the performance target anyhow. These results are stable to tests for seasonality and time trends and other robustness tests.

2 Data

2.1 The use of personnel data for economic research

Over the last decades, personnel data have been increasingly used for research in labour economics (see Ichniowski and Shaw (2012) and the references cited therein). Personnel data mostly contain information on individual workers of a single firm or information on comparable establishments from one sector. An attractive feature of most personnel data is that they contain direct measures of performance. Examples of direct measures of performance are, at the establishment level, quality-related measures such as the establishment's scrap rate (Holzer *et al.*, 1993) or efficiency-related measures such as the uptime of production lines (Ichniowski *et al.*, 1997).

Personnel data have various advantages for economic research. First, most personnel data have the advantage of having direct performance measures at the individual (worker) level, which is typically unavailable in representative surveys across occupations. These performance measures are as widespread as the types of firms analysed in personnel data studies. These output measures range from the number of windshields installed (Lazear, 2000) to the number of trees planted (Shearer, 2004) or the number of fruits picked (Bandiera *et al.*, 2005).

Second, compared to large-scale survey data that have the advantage of providing a representative sample of one or more sectors of industry, personnel data mostly include all workers of one firm or a firm's establishment or department. Personnel datasets are therefore inherently representative of the underlying population but come at the cost of external validity, that is, the generalisability of the results from personnel data to a wider working population.

A third advantage of personnel data over large-scale survey data is that they often contain very detailed firm information, which may include information on human resource management practises applied (see, e.g., Ichniowski *et al.*, 1997). This reduces the problems of unobservable heterogeneity between firms (Ichniowski and Shaw, 2012).

A fourth advantage of using personnel data is that they often allow the researcher to exploit changes induced by a firm—for example, the introduction of new human resource management practices such as incentive pay (see Chapter 6)—or to introduce exogenous variations by running field experiments within the firm (see Chapter 4).

These key advantages of personnel data often allow the researcher to obtain new insights into economic issues that can only be handled inadequately through using (representative) survey data. Personnel data can therefore complement traditional data and offer new insights.

2.2 Data used in this thesis

Chapters 3 through 6 build on data collected from the call centre of a multi-national telecommunications company in the Netherlands. These data contain rich information on individual worker performance for any of the workers' working weeks. The data thus comprise a large panel of performance information at the individual (worker) level. Throughout the thesis, I use data in which the individual observation is the worker–week, that is, the performance of worker i in week t .

Apart from the rich information on performance outcomes, the data have two key characteristics: First, an experimental intervention introduced a training programme to which workers were randomly assigned. This randomisation is key to the identification strategy used in Chapter 4 to estimate the causal effects of training participation on performance outcomes. Second, firm management introduced performance bonuses that depend on predefined performance targets. This discontinuity is used to identify the effects of performance pay on performance outcomes in Chapter 6.

2.2.1 Organisation of the firm

The call agents included in the data work on inbound customer calls only. Customers call the company and are then connected to one of the available agents in the call centre. All agents have the same task of handling these customer calls. Other interactions with customers, such as written correspondence, are dealt with by other back office employees.

Agents are organised into teams that are supervised and monitored by a team leader. The team leaders are supervised and monitored by the call centre's general manager. In general, all teams provide all services; that is, there is no specialisation of teams to handle certain types of calls or further specialisation for certain types of customers. Work places are organised into work islands, with up to eight agents on a team sitting next to each other. Although there is no team work involved in the production process, the spatial proximity implies that there is a scope for peer effects through learning from peers by exchanging information on the work floor. This spatial proximity is exploited to identify peer effects in informal (Chapter 3) and formal (Chapter 4) learning.

Explicit incentives in the firm consist of two parts: annual wage increases and bonus payments by the firm to the individual worker. Annual wages depend on a formal evaluation by the team leader. Annually, in April or May, agents are evaluated

2 Data

and receive a grade from one (lowest) to five (highest). This grade is then used as a multiplier for the reference wage increase. If management sets the reference wage increase at 4%, a grade 1 agent receives no wage increase, grade 3 agents receive 4%, and agents with the highest grade (five) receive 6% (150% of the reference wage increase). Management's aim is to achieve a normal distribution of grades where agents hardly ever receive grade 1 or grade 5 and most agents receive grades 2 to 4.

The firm's bonus payments were changed in April 2011 (see Chapter 6). Before this date, a formal evaluation by the team leader was used to adjust a reference bonus payment for the individual agent's past performance. Since April 2011, agents are incentivised by receiving specific performance targets on one performance measure.

2.2.2 Performance measures

A huge advantage of the data used in this thesis is that all the agents included in the analysis do the same type of work, handling inbound customer calls, and have no other tasks. This allows me to use available performance measures to compare performance across time (within agents) and across agents (between agents).

Average handling time

Handling time, the performance outcome used in Chapters 3 through 5, was one of the main key performance indicators used by the call centre's management before April 2011, as well as in most other call centres. It is defined as the average time an agent needs to handle a customer call (Liu and Batt, 2007; Breuer *et al.*, 2010). The handling time includes the time needed to talk to the customer and log the call in the customer database. Information on the average handling time of each individual agent i is available for each working week t .

The department's aim is to improve firm performance by decreasing average handling time. Throughout this thesis, I therefore use the inverse of the average handling time (aht_{it}) multiplied by 100: $y_{it} = \frac{1}{aht_{it}} \cdot 100$. Since lower values of aht_{it} are interpreted as higher performance, this transformation allows me to interpret improvements in y_{it} as increases in individual productivity.

Service quality

The bonus introduced in April 2011 is based on a new key performance indicator: service quality. This performance measure, used in Chapter 6 only, is generated from a customer satisfaction survey among randomly chosen customers. Among other ques-

tions, customers are asked whether they would recommend the telecommunications company to their family and friends, based on the previous call. This question can be answered on a scale from zero (‘very unlikely’) to 10 (‘very likely’). From the answers to this question, management calculates the so-called net promoter score (see, e.g., Keiningham *et al.*, 2007), which is assumed to be related to customer loyalty. As the question explicitly asks for the customer’s opinion based on a particular call, I am able to use this as a measure of an agent’s service quality.

Alternative performance outcomes

Besides these two main performance outcomes used to assess agent performance, I employ a number of alternative performance outcomes that are mainly used to assess the service quality provided by agents.

- **Share of repeat calls:** Chapters 3 and 4 use the share of repeat calls as an alternative measure to assess agents’ service quality provided. This measure is defined as the share of customers an agent talked to who called the call centre again within seven days. The firm uses this measure as an indicator of call quality because low customer satisfaction with the first call is the main reason for repeat calls.
- **Survey information:** In addition to information on the share of repeat calls, Chapter 4 uses data gathered in a customer satisfaction survey. Individual calls made by call agents were randomly selected and evaluated by customers on a scale from one (very bad) to 10 (very good). Customers gave grades in three different dimensions: the ‘knowledge of the agent’ (*Grade 1*), whether the ‘agent understood the question’ (*Grade 2*), and whether the agent had a ‘solution to the problem’ (*Grade 3*). Because the week t in which the call was made and the corresponding agent i are known, I am able to match this information with information on whether the agent already followed the training program or not. The unit of observation for this analysis is thus a single call and not agent–week information.
- **Share of problems solved:** Chapter 6 uses the share of problems solved as a second performance outcome to approximate service quality. The information is taken from a customer survey that was only introduced in January 2010.¹

¹This survey is the same as the one that gathers information on service quality. The data were not used in any of the other chapters of this thesis, since they have only been available since 2010.

Customers were asked whether their ‘question was completely resolved’ by the call agent in the corresponding call (yes/no). This variable is defined as the number of solved problems over all evaluations of agent i in week t .

2.3 Comparison to other firms and other sectors

As described in Section 2.1, the use of personnel data has several advantages, such as detailed information on the organisation of the firm and performance measures that are comparable across workers. These advantages, however, come at the cost of representativeness. Compared to (representative) surveys or administrative data, it is more difficult to generalise results as the size of the estimates for other firms will also depend on the type of job and the organisation of the firm.

The call centre sector is a comparably young sector that grew substantially since the 1980s, facilitated by the increasing availability of information and communication technology infrastructure. Batt *et al.* (2005) estimate that call centres in the United States employ roughly 4 million employees, which is about 3% of the total workforce. Most of these call centres are in-house centres (Batt *et al.*, 2009). In the Netherlands, the picture is fairly similar: Using data from the Global Call Centre Survey, Holman *et al.* (2007) show that roughly three quarters of all call centres in the Netherlands are in-house call centres.

There are two major differences between call centres and other firms that are important when generalising the results obtained in this thesis. First, work-related training is very important as there is no specific vocational training for call agents. For this reason, call centres need to invest substantially to train their work force. Compared to jobs requiring a certificate of vocational or higher education, the estimated effects of training on performance as well as the tenure-performance profiles might therefore be different for call agents.

Second, advanced monitoring technology allows call centres to monitor the performance of their agents at very low costs. Key performance indicators are usually extracted from the IT-system. It is thus not surprising that almost 40% of the call agents’ pay in the Netherlands is subject to individual performance (Holman *et al.*, 2007). Management as well as agents can easily see the performance development and are able to compare that it to that of other agents. These low costs of monitoring make it harder for workers to underperform.

3 The role of peers in estimating tenure-performance profiles

3.1 Introduction

Besides investments in formal training, tenure has shown to be an important determinant for workers' productivity as indicated by their wages (Mincer, 1958).¹ The main argument for wages increasing with tenure is that workers accumulate human capital which is then rewarded by higher wages. Previous literature on tenure-performance profiles shows that there are indeed substantial gains from experience (Bishop, 1989; Kostiuk and Follmann, 1989; Shaw and Lazear, 2008). It is, however, not only likely that workers learn from practising their tasks, but also from observing or imitating peers at work (Barron *et al.*, 1989). When there are possibilities for learning from others, a worker's output may not only increase with tenure, but also through being exposed to more experienced peers.

In this chapter, we estimate tenure-performance profiles for the first year in a job using individual performance information of call agents taken from personnel records of a large call centre, which is part of a multi-national telephone company in the Netherlands. We find that there is a large marginal return to tenure in the first months of working for the call centre which decreases substantially after gaining some experience. This profile closely follows a logarithmic specification. A 10% increase in tenure is related to an increase of 5.46% of one standard deviation in performance. For the first 12 months of working in the department analysed, this amounts to an increase of about 75% in performance. We interpret this increase in performance as the result of learning on-the-job. When we relate our findings to data on the number of hours the firm invests in formal training, we find that the number of hours invested in learning on-the-job is almost twice as large as the investments in formal training programmes for newly hired agents.

Furthermore, we analyse peer effects in learning, and estimate whether the shape of the estimated tenure-performance profile depends on the composition of teams. Our findings show that there are peer effects when agents are working with more experienced peers in their team. If average peer tenure increases by one standard deviation, agents' performance increases by 11 to 14% of a standard deviation in performance. Our results suggest that in the workplace we analyse learning-by-doing is more important for learning than learning from others. We show that our results are robust against alternative hypotheses, especially against the hypothesis of selective labour turnover. Furthermore, we show that the number of hours worked affects the

¹This chapter is based on joint work with Andries De Grip and Inge Sieben (De Grip *et al.*, 2011). We would like to thank Eric Bonsang, Ben Kriechel, Olivier Marie, seminar participants at Maastricht University, SOLE (Cambridge, MA), EALE (Tallinn) and EEA (Barcelona) for helpful comments.

shape of the estimated tenure-performance profile; a result which may be explained by ability-sorting. In addition, we test whether the increase in performance, which is measured as the average length of calls, comes at the cost of decreases in quality. Although we do not find a negative effect on direct measures of customer satisfaction, we do find a negative relation between tenure and the number of customers calling back within a few days after the call. Overall, it seems that the increase in quantitative performance outweighs the decrease in the number of customers calling back.

This chapter contributes to the literature on estimating the effects of experience on workers' productivity. These studies originate from the literature on estimating the effects of tenure on wages (Lazear, 1979; Baker *et al.*, 1994; Altonji and Williams, 2005). One of the main arguments for upward sloping tenure-wage profiles is that it reflects an increasing stock of human capital which is rewarded by higher wages (Mincer, 1958). It has been shown, however, that estimating tenure-wage profiles is problematic for a variety of reasons, such as controlling for trends in wages, sample selection and sample attrition (Altonji and Williams, 2005). Furthermore, using wages instead of output may underestimate the actual returns on learning-by-doing (Shaw and Lazear, 2008). One way to overcome these issues is to use performance data from personnel records. Data taken from personnel records often contain direct measures of performance and can deal with sample selection and attrition since the full sample is observed. Carroll *et al.* (1986) and Kostjuk and Follmann (1989) use the number of contracts produced by Navy recruiters as a proxy for performance and find that individual performance substantially increases in the first months of their employment. Shaw and Lazear (2008) use information about performance of windshield installers in a US-American firm to estimate tenure-performance profiles, and find that output increases by 53% within the first eight months on the job.²

In contrast to these previous studies on the estimation of learning curves, we also analyse whether working with experienced peers affects learning. Barron *et al.* (1989) provide evidence that workers spend a substantial part of learning investments on watching other workers performing their tasks. Their findings suggest that the working environment is important for the learning process in the workplace. Destré *et al.* (2008) estimate the relative importance of self-learning and learning from others using establishment data from France. They show that workers can learn about 10% of their own stock of human capital from co-workers. Battu *et al.* (2003)

²There is a related literature analysing learning at the firm level (Benkard, 2000) which distinguishes between experience effects and spill-over effects in learning within firms (Thornton and Thompson, 2001).

3 *The role of peers in estimating tenure-performance profiles*

and Moretti (2004) find evidence for positive human capital spill-overs within firms and across firms, respectively. Several studies using information from personnel records find positive externalities either through social ties (Bandiera *et al.*, 2010) or simply through spatial proximity (Sacerdote, 2001; Zimmerman, 2003; Falk and Ichino, 2006; Mas and Moretti, 2009). In contrast to these studies, Guryan *et al.* (2009) do not find significant peer effects on workers' performance, suggesting that peer effects might be related to specific industries, occupational groups or tasks.

The data used in this chapter have several key features that allow us to estimate unbiased tenure-performance profiles and the effect of peers on the shape of this profile. First, we observe *all* agents working in the call centre over a period of 30 months. All agents must perform the same tasks irrespective of their tenure, and face the same incentives set by the call centre's management. Given the relatively high turnover of agents, this setting provides substantial variation in tenure and a large number of newly hired agents, which allows us to estimate tenure-performance profiles across tenure in the first year in the job. Furthermore, we are able to control for endogenous selection of call agents out of the call centre.

Second, our data provide very detailed information about individual performance, observable for each week an agent works. Following other studies using similar data on call centre agents (e.g. Castilla, 2005; Liu and Batt, 2007), we use a measure of performance that is based on the average handling time of calls. This measure has the advantage of being objective, as it is automatically generated and stored and thus not subject to potentially biased evaluations by supervisors (Flabbi and Ichino, 2001). Furthermore, it is preferable to wage data because it directly measures performance, while wages are often fixed under hourly pay (Shaw and Lazear, 2008). Besides performance information, the data also contain information on investments in formal training programs. As there are no vocational training programmes for call agents, we can calculate workers' total stock of formal training.

Third, the data allow us to analyse peer effects between agents who work in the same team. Despite a growing interest in the effects of social interaction in the workplace, peer effects on worker productivity are difficult to identify (Manski, 1993). A few studies use truly randomised variation in the assignment of workers to peers to identify peer effects (e.g. Chapter 4 of this thesis). A second strand of studies use quasi-exogenous variation of workers in teams where one would not expect endogenous sorting (e.g. Mas and Moretti, 2009). In this chapter, agents who belong to the same team are located next to each other on the work floor. We exploit the quasi-random variation in assignment to shifts as well as to teams to identify peer effects on workers' individual performance.

The chapter is structured as follows. In Section 3.2, we describe the setting of this chapter and present arguments for both learning-by-doing and learning from others. In Section 3.3, we discuss our empirical model and present the estimates of the tenure-performance profile of call agents and the effects of peers on workers' tenure-performance profile. Section 3.4 shows some robustness analyses and Section 3.5 summarises and concludes.

3.2 The firm

3.2.1 Information on the workplace

To estimate tenure-performance profiles and the impact of peers on the shape of these profiles, we use panel data of call agents in a call centre. The call centre belongs to a multi-national telecommunications company in the Netherlands, and is segmented in five departments which are distinguished by the type of customer. In order to analyse a homogenous production process, we focus our analysis on the largest department which handles inbound calls of private customers with fixed contracts. In the course of the observation period, 440 agents worked for this department. All agents have the same task of handling customer calls in case of questions, complaints or problems. Other interactions with customers such as written correspondence are dealt with by other back-office employees.

The data contain objective performance information for each call agent in each week that the agent is working. The performance measure y_{it} is based on the average handling time aht_{it} of agent i in calendar week t . It measures the average time an agent spent talking to customers and logging the information about the call in the customer database. This measure has the advantage in that it is automatically generated and is not affected by potentially subjective performance evaluations by supervisors. In recent studies using data on call agents, similar outcomes are used to proxy an agent's productivity (e.g. Castilla, 2005; Liu and Batt, 2007; Chapter 4 of this thesis). Furthermore, it is used by call centre management as the key performance indicator for monitoring performance of agents in the call centre. The main objective of management is to decrease average handling time without decreasing the quality provided by the agents. We therefore interpret shorter calls as beneficial to the firm, and define productivity as $y_{it} = \frac{100}{aht_{it}}$. Decreasing average handling time aht_{it} is thus interpreted as an increase in performance y_{it} .³

³We deal with potential quality-quantity trade-offs in Section 3.4.3.

3 The role of peers in estimating tenure-performance profiles

Agents are organised into 16 teams, each of which is led by a team leader who is responsible for monitoring and coaching agents. In general, all teams provide all services, that is, there is no specialisation of teams in handling certain types of calls or serving certain types of customers. All teams work on the same floor of the building. Work places are organised into work islands, with up to eight agents of a team sitting next to each other. Though there is no team work embedded in the production process, the spatial proximity implies that there is scope for peer effects through learning from peers by exchanging information.

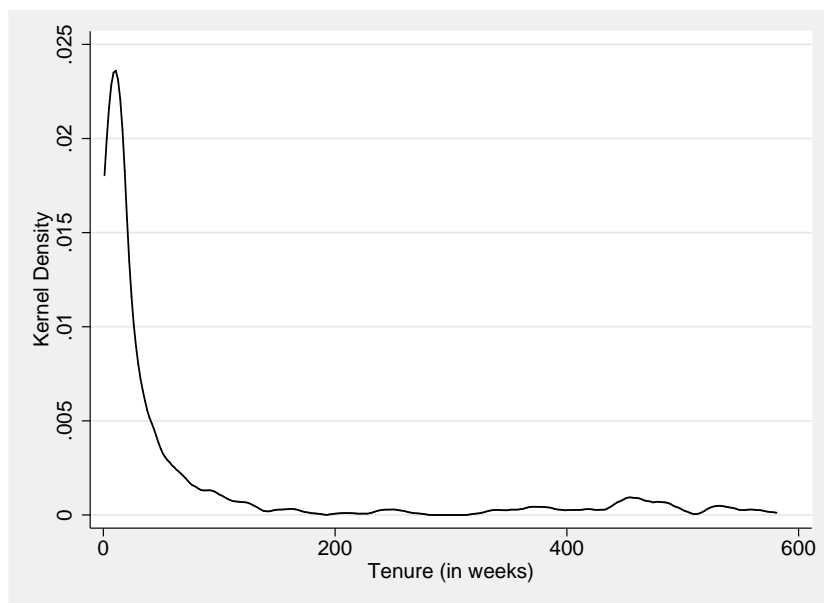
Although agents' performance is continuously measured, agents are paid a flat pay with an adjustment once a year. Agents can receive a wage increase of up to 8% depending on an annual performance rating by the supervisor.

3.2.2 Sample selection and descriptive statistics

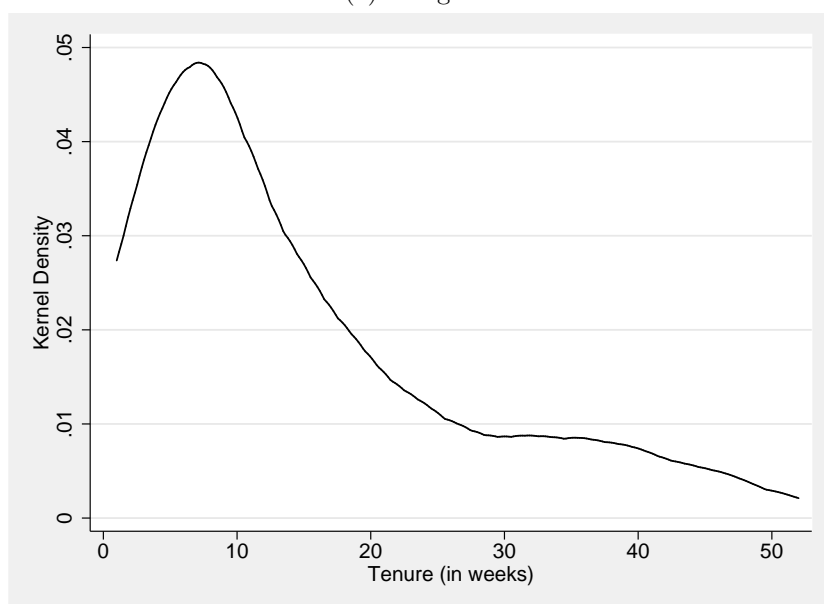
Our data covers all agents working in the selected department between October 2007 and March 2010. In total, 440 individuals were working for this department for at least one week in this period. To estimate tenure-performance profiles of agents, however, we restrict the sample to all employees with less than one year of tenure. There are several reasons for this restriction. First, focusing on the first year of tenure reduces the likelihood of biased results due to selection because personnel turnover is relatively high in call centres in general (Batt *et al.*, 2005). In this call centre, 49% of all agents observed in our data leave the department within the first year, either to other departments within the call centre or out of the firm. Because this selection is potentially non-random, we focus on the first year of tenure to limit the effect of selection on the estimated tenure-performance profile. The high turnover of call agents also implies that there are less data to estimate tenure-performance profiles beyond one year of tenure (Panel (a) of Figure 3.1), while there are sufficient data for estimating tenure-performance profiles among agents in the first year (Panel (b) of Figure 3.1).⁴

⁴To reduce the number of outliers, we also drop all observations that have calls shorter than the first percentile or longer than the ninety-ninth percentile. In most cases, these extreme average handling times were caused by a very low number of calls.

Figure 3.1: Kernel density plot for tenure



(a) all agents



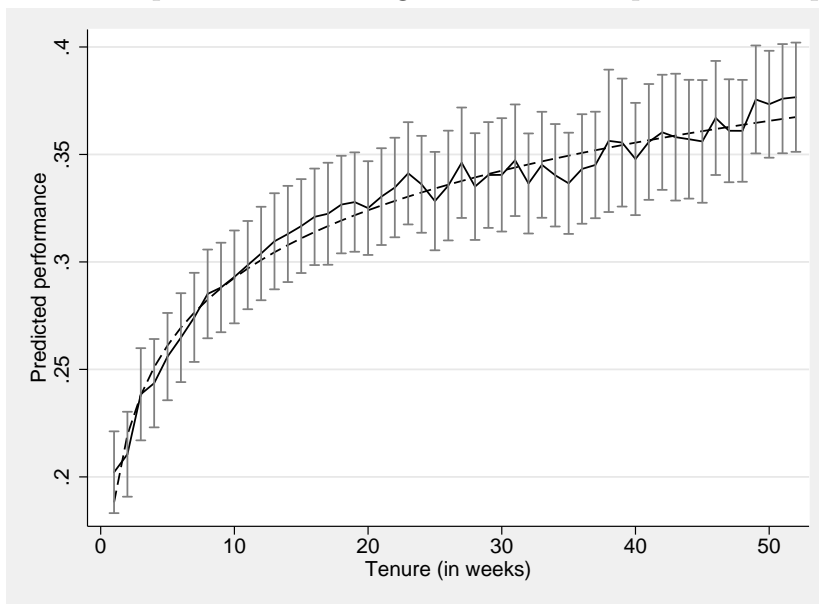
(b) newly hired agents

Note: Panel (a) shows the the density plot for all agents, (b) shows the density plot for all agents with a maximum of one year of tenure. Tenure is measured in weeks and averaged for each agent in the sample.

3 The role of peers in estimating tenure-performance profiles

Second, human capital theory implies that learning effects are highest in the beginning of an agent's career. Because of faster learning, it is reasonable to expect that workers' performance is more likely to be affected by tenure in this period compared to later periods. This pattern can easily be seen by the solid line depicted in Figure 3.2 which shows that the average performance \bar{y}_{it} increases in tenure. After a very steep increase in the beginning, the marginal returns to tenure decrease. This figure shows that there is substantial learning in the agents' task in the first months of the employment relationship, but much less thereafter.

Figure 3.2: Non-parametric and logarithmic tenure-performance profiles



Note: the solid line shows the predicted performance after a regression of performance y_{it} on tenure fixed-effects and week fixed effects. The capped spikes show the appropriate 95% confidence interval. The dashed line shows the predicted performance after a regression of y_{it} on $\log(\text{tenure})$.

Column (1) of Table 3.1 shows the descriptive statistics of all agents working in the call centre department we analyse ($N = 440$). In the sample, 33% of the agents are men, and their average age is 29 years (Column (1) of Table 3.1). The average number of working hours is 21.4 hours per week. These figures are comparable to call centres in general which are characterised by a predominantly female workforce with a substantial proportion of part-time workers (Batt *et al.*, 2009).

Table 3.1: Descriptive statistics (main variables)

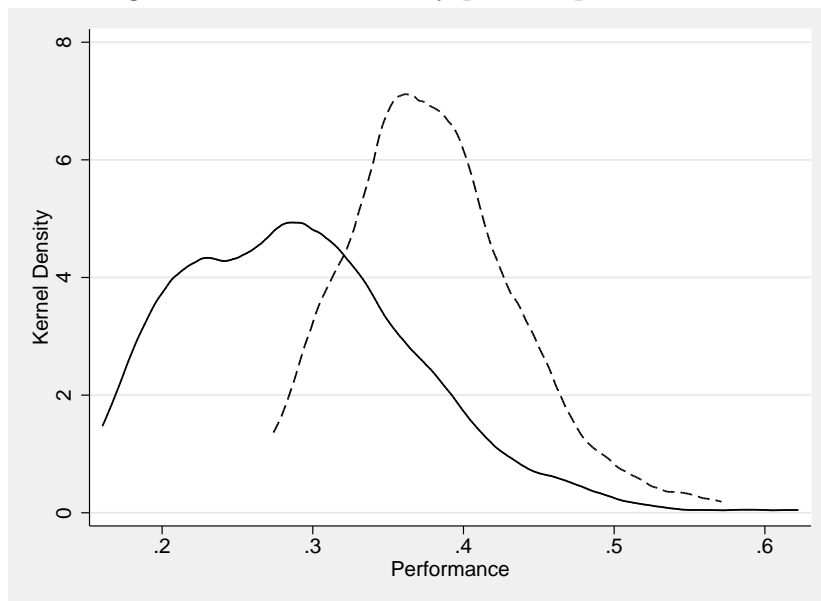
	(1)	(2)	(3)	(4)	(5)	(6)
Tenure restriction	No	Up to 1 year of tenure				
Sample restriction	All agents	All agents	New hires	Leavers	Stayers	Diff. (5)-(4)
Performance (y_{it})	0.314 (0.088)	0.297 (0.081)	0.283 (0.072)	0.279 (0.070)	0.292 (0.075)	0.016*
Tenure (in years)	1.360 (2.606)	0.297 (0.232)	0.227 (0.153)	0.221 (0.152)	0.241 (0.155)	0.072**
Previous employment	0.112 (0.316)	0.045 (0.207)	0.013 (0.114)	0.005 (0.070)	0.031 (0.175)	0.026*
Working hours	21.433 (7.489)	23.119 (7.121)	23.647 (7.298)	21.189 (5.549)	28.945 (7.812)	7.784***
Gender (1=male)	0.328 (0.470)	0.351 (0.478)	0.356 (0.480)	0.348 (0.477)	0.375 (0.487)	0.027
Age	29.331 (10.105)	27.011 (8.192)	26.173 (7.846)	26.506 (7.496)	25.525 (8.492)	-0.981
Number of agents	440	356	303	207	96	
Number of observations	15011	7623	6451	4274	2177	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Descriptive statistics are based on the average for each individual agent. Standard deviations in parenthesis.

3 The role of peers in estimating tenure-performance profiles

In Column (2) of Table 3.1, descriptive statistics of the estimation sample, which comprises all observations with less than one year of tenure, are shown. Compared to the full sample of workers (Column (1)), agents with less than one year of tenure have on average lower performance (0.297 compared to 0.314). Figure 3.3 shows both the kernel density of performance for all agents (dashed line) and for agents with a maximum of one year of tenure (solid line). It illustrates that there is a substantial shift from the performance of starting agents to the performance of more experienced agents.

Figure 3.3: Kernel density plot for performance



Note: The figure shows the kernel density for performance y_{it} for all agents with a maximum of one year of tenure (solid line) and for all agents with more than one year of tenure (dashed line). Performance is measured in years and averaged for each agent in the sample

To shed more light on non-random selection of agents out of the sample, we focus on agents which we can observe from their first tenure-week (Column (3) of Table 3.1). Within this sample, we can precisely observe when and which agents leave the sample (see Shaw and Lazear, 2008). When comparing the observations of agents leaving the sample within the observation period ('leavers', Column (4)) with those staying in the sample during that period ('stayers', Column (5)), we find that stayers are slightly more productive, have more tenure, and work substantially more hours than leavers (Column (6)). Though the group of leavers can comprise positive as well as negative sorting of agents, this result suggests that agents leaving the department are on average a negative selection.

3.3 Estimating tenure-performance profiles

3.3.1 Baseline results

The average tenure-performance profile for agents in their first year of working for the department is shown in Figure 3.2. The solid line depicts average performance, controlling for general time-trends and fluctuations which are measured by week fixed-effects. The figure shows a steep tenure-performance profile in the first six months of working in the call centre. After this period, performance is still increasing, but at substantially lower rates than in the first six months. This result suggests that there is substantial learning about how to handle customer calls in the first months of the employment relationship but not thereafter.

In order to estimate the agents' tenure-performance profile, we model the effect of tenure d_{it} on performance y_{it} as a linear-log function. The linear-log prediction, which is depicted by the dashed line in Figure 3.2, provides the best approximation of the non-parametrically estimated tenure-performance profile compared to other functional forms. Our regression model can thus be written as:

$$y_{it} = \alpha + \log(d_{it})' \beta_1 + X_{it}' \beta_2 + \gamma_t + u_{it} \quad (3.1)$$

where the unit of observation is agent i in calendar week t . The slope of the tenure-performance profile is thus estimated by β_1 . The vector X_{it} indicates whether the agent had a previous employment relationship with the same call centre, the number of working hours in week t , as well as the agent's gender. Average performance fluctuates substantially around its mean, due to such issues as technical network problems, problems with the internal IT-system, or changes in the composition of types of calls. To control for these aggregate shocks, we add week fixed-effects γ_t to Equation (3.1). The effect of tenure on performance can be identified from general time trends because the timing of entry dates varies across all agents. The idiosyncratic error term u_{it} is clustered to allow for within-agent correlation.

In general, it is possible to estimate Equation (3.1) including individual fixed-effects (see Shaw and Lazear, 2008). Because tenure and time are perfectly collinear, however, the effect of tenure cannot be identified unless making ad-hoc assumptions

3 The role of peers in estimating tenure-performance profiles

such as the equality of two specific fixed effects.⁵ Equation (3.1) is therefore estimated using pooled ordinary least squares.

The main regression results are shown in Table 3.2. Despite piecewise adding control variables X_{it} and week fixed-effects to the regression, the estimated effect of tenure on performance ($\hat{\beta}_1$) is hardly affected and stays at about 0.045. Our preferred specification, which is shown in Column (5), includes the full set of covariates. The results show that, over the first year working for the call centre, an increase of 10% in tenure is related to an increase in performance y_{it} of 0.00442. Given a standard deviation of performance of 0.081, this relates to an increase of 5.5% of a standard deviation in performance.

Table 3.2: The effect of tenure on performance

	(1)	(2)	(3)	(4)	(5)
log(tenure)	0.0457*** (0.0024)	0.0448*** (0.0025)	0.0446*** (0.0024)	0.0448*** (0.0024)	0.0442*** (0.0026)
Previous employment		0.0415*** (0.0134)	0.0399*** (0.0137)	0.0411*** (0.0138)	0.0399*** (0.0138)
Working hours			-.0003 (0.0003)	-.0003 (0.0003)	0.0001 (0.0003)
Gender (1=male)				0.0109 (0.0089)	0.0111 (0.0088)
Constant	0.3799*** (0.0066)	0.3761*** (0.0069)	0.3830*** (0.0108)	0.3795*** (0.0109)	0.3690*** (0.0116)
Week fixed-effects	No	No	No	No	Yes
Observations	7623	7623	7623	7623	7623
Number of agents	356	356	356	356	356
R ²	0.2124	0.2224	0.2234	0.2261	0.2918

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: y_{it} . All regressions are estimated using observations with a maximum of one year of tenure. Standard errors are clustered at the agent level.

To control for agent-specific characteristics, we included a set of control variables. First, to control for previous employment spells as a call agent, we use a dummy variable previous Employment which is defined as being one (1) if the agent had a employment relationship with the same call centre before the current employment relationship. Among agents in our estimation sample, 4.5% had previous employ-

⁵In order to assess the impact of individual fixed effects on the estimated tenure-performance profile, we also estimated Equation (3.1) with and without individual fixed effects, both without week fixed effects. The results show that there is no statistical difference between the effect of the tenure variable of the two models. The results are available upon request from the authors.

ment relationships with this call centre (Column (2) of Table 3.1). Agents who had an employment relationship with the call centre prior to the current employment relationship perform 0.0399 (i.e. 49.3% of one standard deviation in performance y_{it}) better than agents without previous employment. This suggests that agents who previously worked in this centre have accumulated skills which enable them to work at a faster pace, compared to new hires without previous employment in the call centre.

Second, we include the number of weekly working hours in order to control for the amount of time an agent spends on learning-by-doing. Third, we control for an agent's gender. Both variables, however, do not have a significant effect on the shape of the estimated tenure-performance profile.

3.3.2 Investments in learning on-the-job

Our interpretation of the shape of the tenure-performance profile is that increases in performance reflect an agent's stock of skills. The shape can thus be used to estimate the size of the firm's implicit investment in learning on-the-job, G . Given supply of experienced workers, the firm could either appoint an experienced agent for a vacant position, or can hire an inexperienced agent with zero tenure to fill the vacancy. The implicit investment comes from the fact that inexperienced agents are learning and thus perform worse than experienced agents in the first months of employment. We therefore assume that an inexperienced agent starts with performance $y_{d=1,it}$ in the first working week $d = 1$ and then follows the tenure-performance profile estimated by Equation (3.1) until the agent can work at the performance level of an experienced agent in (tenure) week $d = K$, $y_{d=K,it}$.

After calculating the difference between the two agents $y_{d=K} - y_k \forall k = 1, \dots, K-1$, we transform this number to the difference in average handling time, and multiply it by the average number of calls an experienced agent handles in one week, $c_{k,it}$. We interpret the difference, G ,

$$G = \sum_{k=1}^K \left(\frac{100}{y_{d=K} - y_k} \right) \cdot c_{k,it} \quad (3.2)$$

as the learning investment of untrained workers. When estimating G for an average agent, we find that 161.7 hours of learning are required before a new hire has the same productivity as an experience agent. All other things equal, the firm has to invest this amount of 'learning on-the-job', minus the wage difference between the two agents. In fact, these opportunity costs of hiring an inexperienced worker are

about twice as high as the call centre’s investment in hours of formal training, which accumulates to 84.1 hours of formal training programs in the first year.

3.3.3 Team composition and tenure-performance profiles

The actual task of agents, handling customer calls, is an individual production process.⁶ Despite the fact that there is no joint team work, peer effects are plausible in our context, because agents do the same task regardless of their tenure, and teams are located next to each other at the workplace. Peer effects are more likely between agents of the same team, compared to peer effects between agents of different teams, because agents of one team sit next to each other, while agents of other teams have a larger spatial distance at the work floor. Moreover, agents who belong to the same team may exchange work-related information and knowledge during team meetings.⁷ In addition, agents with higher tenure probably have more skills to handle calls, retrieve information from the customer to deal with the customer’s problem efficiently, and to use the computer software system, compared to new hires. This argument is supported by the outcomes of the analysis in Section 3.3.1, which shows a steep tenure-performance profile over the first year of tenure.

We use three different variables to capture the effect of peers on a worker’s own performance. The first two variables used to proxy peer effects are average performance of peers in an agent’s team, and the peers’ average tenure. We use both variables lagged by one period, i.e., one working week. The reasoning for using lagged values instead of contemporaneous values is twofold. First, one argument for peer effects is that workers learn from their peers. It is more likely that peer performance and peer tenure do not have an immediate effect but should rather pay off *after* having learned from peers. Second, using lagged peer performance avoids the problem of simultaneous determination of both outcome y_{it} (agent’s performance) and peers’ performance, for instance due to common shocks to performance.

The third variable used to proxy peer effects is the average peer experience, cumulated over an agent’s tenure, and normalised by the agent’s tenure. The reasoning

⁶This does not preclude that there are externalities from an individual worker’s production to the worker’s peers. In the present setting, providing low quality can lead to customers calling back. These repeat calls may have to be treated more carefully. We argue, however, that this does not affect our estimate. First, agents are monitored thoroughly by their team leaders who can listen in on calls. Second, since calls are randomly assigned to agents in the call centre, all agents have the same probability of receiving calls from customers calling back.

⁷This argument does not preclude that peer effects also arise between agents of different teams. If there are also between-team peer effects, our estimated within-team peer effect can be interpreted as a lower bound of the true peer effect.

Table 3.3: Descriptive statistics (other variables)

	(1)	(2)	(3)	(4)
	mean	SD	Observations	Agents
Average team performance $t - 1$	-0.207	0.857	3194	227
Average team tenure $t - 1$	-0.379	0.901	3155	227
Average peer experience / tenure	-0.502	0.893	3430	237
Permanent productivity ^a	0.192	0.064	345	171
Call quality (y_{it}^q)	0.507	0.178	7489	355
Composite productivity (y'_{it})	0.249	0.074	7489	355
Grade 1	6.876	2.232	841	112
Grade 2	6.404	2.294	850	112
Grade 3	6.034	2.734	776	105
Tenure in working weeks (in years)	0.183	0.145	6451	303
Tenure in fte weeks (in years)	0.317	0.239	6451	303

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Descriptive statistics are based on the average for each individual agent. Agents with one year of tenure at most are considered. Standard deviations in parenthesis. Permanent productivity is calculated from fixed effects θ_i which are obtained from a regression of average handling time (y_{it}) on individual fixed effects, and a full set of dummies for each level of tenure d (see Equation 3.4). For the calculation of the variable composite productivity, see Section 3.4.3. Grades are given on ‘knowledge of agent’ (grade 1), ‘agent understood question’ (grade 2), ‘solution of the problem’ (grade 3). Grades are given on a scale from 1 (very bad) to 10 (very good). Tenure in working weeks is defined as the number of working weeks an agent worked. Tenure in fte weeks is defined as the number of fte working weeks an agent worked.

^a based on the full data.

for this variable is that we want to measure the cumulative effect of being exposed to more experienced agents; the variable therefore measures the cumulative exposure to tenure. To be able to compare the effects of the three peer variables, all are standardised with a zero mean and a standard deviation of one. Descriptive statistics of the peer variables are shown in Table 3.3.

The effect of the peer variables on workers’ performance is estimated in three separate regressions by augmenting Equation (3.1) by the variable d_{it} which captures information on peers:

$$y_{it} = \alpha + \log(d_{it})' \beta_1 + X'_{it} \beta_2 + \gamma_t + d'_{it} \theta + u_{it} \quad (3.3)$$

We estimate the effect of the three distinct definitions of the variable d_{it} separately. Each of the peer variables is calculated as the average for agent i in the previous week $t - 1$. All peer variables take into account that an agent is exposed to another agent who works the same hours because the peer variables are weighted by the number

3 The role of peers in estimating tenure-performance profiles

of joint working hours of agent i with the worker's peers' working hours in the same week.

Because information on when an agent enters and leaves the call centre is available for a shorter time period only, the number of observations drops from 7,623 (356 agents) to 3,430 (237 agents). For this reason, Column (1) of Table 3.4 first shows the estimation results of Equation (3.1) with the sample used to estimate peer effects. The estimated effect of log-tenure on the measure of performance, y_{it} (0.048) does not significantly differ from the previous estimate (Column (5) of Table 3.2). The other columns of Table 3.4 shows the estimation results of Equation 3.3. Columns (2) and (3) show the effect of peer performance and peer tenure on a worker's own performance, whereas Column (4) shows the effect of cumulated peer experience. All three variables significantly affect workers' own performance, with the main effect of tenure not significantly changing from the baseline model (Column (1)). When including the average performance of agents who were working with the agent in the previous week (Average team performance $t - 1$), log-tenure has an effect of 0.049 (Column (2)). An increase of one standard deviation in average team performance in the previous week ($t - 1$), is related to an increase in workers' own performance of 0.0097. This is the equivalent to 11% of one standard deviation in the outcome variable y_{it} . The use of peer performance may, however, raise concerns, because it may be driven by simultaneous determination of both workers' own performance y_{it} and peer performance d_{it} . We therefore estimate Equation (3.3) with lagged average team tenure as the proxy for peer effects (Column (3) of Table 3.4), and find a significant coefficient of 0.109. This translates to an effect of 13.5% of a standard deviation in y_{it} when increasing average tenure of a worker's team mates by one standard deviation. As a third proxy for peer effects, we use cumulative average team tenure, divided by the number of weeks an agent has worked in the call centre. The estimated coefficient is significant, and relates to an increase in y_{it} of 10.9% of a standard deviation in performance when the cumulative team mates' tenure increases by one standard deviation.

Independent of the specification of the peer variable, the results presented in Table 3.4 show that having worked with better performing or more experienced peers improves a worker's performance. One standard deviation increase in the average performance or the average tenure of a worker's peers leads to an increase in the agent's performance between 11 and 13.5% of a standard deviation. These results show that it may be beneficial to the firm to place new agents in more experienced teams. When calculating G as in Equation (3.2) based on the sample used in Table 3.4, the learning investment accumulates to 164 hours over the first year of tenure.

Table 3.4: The effect of peers' experience on individual performance

	(1)	(2)	(3)	(4)
log(tenure)	0.0476*** (0.0028)	0.0487*** (0.0037)	0.0469*** (0.0041)	0.0449*** (0.0033)
Average team performance $t - 1$ (<i>standardised</i>)		0.0097** (0.0039)		
Average team tenure $t - 1$ (<i>standardised</i>)			0.0109*** (0.0040)	
Average peer experience / tenure (<i>standardised</i>)				0.0088* (0.0053)
Constant	0.3625*** (0.0170)	0.3613*** (0.0173)	0.3605*** (0.0180)	0.3604*** (0.0170)
Week fixed-effects	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes
Observations	3430	3194	3155	3430
Number of agents	237	227	227	237
R ²	0.3945	0.3734	0.3767	0.3991

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: y_{it} . Average team performance is defined as the performance average of peers within the team. Average team tenure is defined accordingly. Average peer experience is calculated as the sum over average team tenure in the past, normalised by the number of periods an agent has been working. All peer variables are weighted by joint hours between agent i and other agents j and standardised with a mean of 0 and a standard deviation of 1. All regressions are estimated using observations with a maximum of one year of tenure. Control variables are previous employment, working hours, and gender. Standard errors are clustered at the agent level.

When including average peer tenure as an additional control variable, this number reduces to 157 hours. If we interpret the difference between the two estimates as the size of overestimation when not correcting for peer effects, the investments G which are due to learning-by-doing are overestimated by 4.3%.

3.4 Robustness analyses

3.4.1 Turnover and selection of workers

Call centres face relatively high turnover of workers compared to other sectors of industries (Batt *et al.*, 2005). If turnover is related to individual performance or to unobserved factors that affect both tenure and performance of the individual worker, the shape of the estimated tenure-performance profile is likely to be affected by workers who quit the firm. Therefore, the observed increase in performance by

3 The role of peers in estimating tenure-performance profiles

Table 3.5: The effects of tenure on new hires

	All (1)	Stayers (2)	Leavers (3)
log(tenure)	0.0444*** (0.0027)	0.0516*** (0.0040)	0.0385*** (0.0034)
Working hours	0.0002 (0.0004)	-.0001 (0.0005)	0.0005 (0.0004)
Gender (1=male)	0.0015 (0.0085)	0.0013 (0.0121)	0.0031 (0.0106)
Constant	0.3685*** (0.0125)	0.4000*** (0.0205)	0.3468*** (0.0141)
Week fixed-effects	Yes	Yes	Yes
Observations	6451	2177	4274
Number of agents	303	96	207
R ²	0.2916	0.3967	0.2610

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: y_{it} . All regressions are estimated using observations with a maximum of one year of tenure. Furthermore, only new hires are included. New hires are defined as agents who are observed from their first working week. Leavers (stayers) are defined as agents who do (not) drop out within our observation period. Standard errors are clustered at the agent level. Previous experience is included from these regressions because of missing variation in one of the regressions.

tenure is partly the result of an improved match quality instead of learning on-the-job.

The sample used in this chapter allows us to compare different types of agents. We can distinguish between agents leaving the department analysed in this chapter (leavers) and agents who are not leaving (stayers).⁸ However, since we do not observe agents who left the department before our observation period started, this analysis is likely to be affected by left-censoring (Shaw and Lazear, 2008). We therefore limit the sample to new hires, i.e. agents that we observe from their first working week.

Column (1) of Table 3.5 shows the results when estimating Equation (3.1) for new hires only. This sample contains 303 agents, of which 96 are stayers; the remaining 207 agents leave within the observation period. The estimated coefficient for log(tenure) is very similar to the baseline estimates (Column (5) of Table 3.2). An increase of tenure by 10% leads to an increase of 5.5% of a standard deviation of performance (estimated coefficient of 0.0444). The estimated effect of log(tenure) on performance for stayers (0.0516) corresponds to an increase in performance of 6.4% of one standard deviation when tenure increases by 10%. For agents leaving

⁸We do not have information on the exact date when agents leave the department. We define agents as leavers when they have not worked in the department for 20 or more weeks in a row.

the department, however, the corresponding effect is only 4.8% (0.0385), which is significantly lower than the estimated performance increase for agents who remain employed in the call centre.

These estimation results have two implications. First, the data shows that a substantial share of agents who are observed from their first working week leave the department within our observation period (68%). Second, these agents have significantly lower tenure-performance profiles. This endogenous sorting of agents can be interpreted as a positive selection of agents who stay employed in the department, while the leaving agents comprise a negative selection of workers.⁹

3.4.2 Assignment to teams and shifts

To identify the causal peer effect on workers' performance development, either truly random assignment by means of experiments (e.g. Chapter 4 of this thesis) or quasi-random assignment to groups (e.g. Mas and Moretti, 2009) is required. In this chapter, the effect of an agent's peers on own performance is identified by quasi-random changes in team composition. Although each agent is assigned to one team supervised by one team leader, there are week-to-week differences in whether an agent is working at all, and in the number of hours an agent is working. Therefore, the agent's composition of team mates differs from week to week.

In order to support the hypothesis of random assignment of agents to teams, we test whether the assignment to teams is selective with respect to ability. We therefore regress a measure of individual ability on the average ability of the team. We generate a measure of ability by using the estimated individual fixed-effects from a regression of performance y_{it} on a full set of tenure and individual fixed-effects (cf. Mas and Moretti, 2009).

$$y_{it} = \eta_i + \gamma_d + \varepsilon_{it} \quad (3.4)$$

This measure of ability, $\hat{\eta}_i$, is then regressed on the average ability of the other workers in the team (Sacerdote, 2001). Because we are interested in the team coaches' influence on the selection of new agents, we run this test for agents switching to new team coaches. The explanatory variable, average peer ability \bar{y}_{ij} , is thus taken from

⁹The difference between the estimated tenure-performance profile of leavers and stayers can be affected by the tenure-segment one observes. When estimating the same regressions for agents with a maximum tenure of 3 months, the estimated tenure-performance profiles are not significantly different from one another. When increasing the tenure-segment, this difference becomes significant.

Table 3.6: Selection into teams

	(1)
Average Permanent Productivity	-.2922 (0.1836)
Constant	0.2504*** (0.0365)
Observations	343
Number of agents	170
R ²	0.005

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: permanent productivity of agent i when switching from team j to team k , y_{ij} . Average permanent productivity is the mean of permanent productivity of all agents in destination team k to which an agent is switching to, \bar{y}_{ik} (excluding the agent's own permanent productivity). Permanent productivity is calculated from fixed effects θ_{ij} which are obtained from a regression of average handling time (y_{it}) on individual fixed effects, and a full set of dummies for each level of tenure d (see Equation 3.4).

the agents in the team j to which an agents is switching. Table 3.6 shows the estimation results of individual ability on average ability in the destination team using agent-week observations after an agent switched to a new team. The results show that average team ability does not have a significant effect on individual ability. This allows us to conclude that there is no ability-based selection for agents when switching team leaders.¹⁰

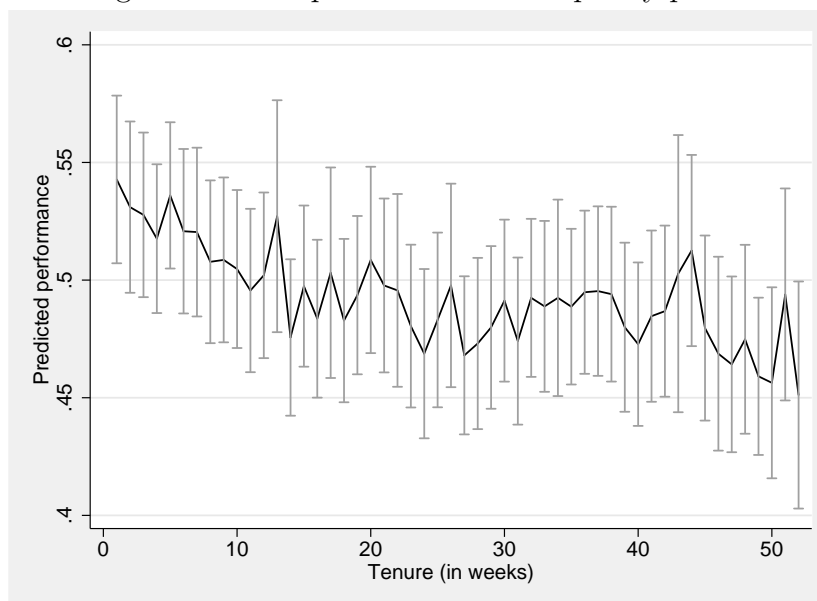
Further, it is rather difficult for agents to select themselves into the same hours as particular peers. Agents must be available for scheduling during the opening hours of the call centre, and must state their preferences for particular hours four weeks prior. Based on both their preferences and the expected demand by customers, managers assign agents to particular hours. As a result, both the composition of agents at each point in time and weekly team composition changes frequently. As the scheduling is done for the whole department and not for individual teams, it is unlikely that there is any strategic planning within teams. Therefore, there is no reason to expect that spurious correlation drives our estimation results of the peer effects.

3.4.3 Substitution between quality and quantity of calls

The performance variable used in this chapter, average handling time, is a measure of quantitative performance. Following Chapter 4 of this thesis, we use additional information to estimate the effect on quality of calls. We employ two types of measures

¹⁰New agents are recruited by a temporary help agency that selects potential agents who fulfill the job description. The team leader usually does not have any influence on the selection of newly recruited agents and their placement into teams.

Figure 3.4: Non-parametric tenure-quality profile



Note: the solid line shows the predicted qualitative performance measure after a regression of performance y_{it}^c on tenure fixed-effects and week fixed effects. The capped spikes show the appropriate 95% confidence interval.

to capture quality of agents' calls. The first measure is based on the share of repeat calls. The share of repeat calls is defined as the share of customers to whom an agent spoke who called the call centre again within seven days. This measure is used by management to evaluate the quality of calls, as customers may call again, potentially because the problem is not solved. Low values of repeat calls (rc_{it}) are interpreted as high performance. We therefore define call quality as $y_{it}^q = \frac{1}{rc_{it} * 10}$, with an average of 0.507 (see Table 3.3). Figure 3.4 shows the corresponding non-parametrically estimated tenure-performance profile for this measure of call quality. Call quality appears to decrease in the first 3 months and to flatten thereafter. The estimation results show that with increasing tenure, call quality significantly decreases (-0.0193; Table 3.7, Column (1)). An increase in tenure by 10% is thus related to a decrease of call quality by 1.1% of one standard deviation in the quality measure. This suggests that despite the fact that quantity is improving substantially over the course of the first months, this partly comes at the cost of lower call quality.

Having information on both average handling time and call quality also allows us to create a composite measure of productivity that contains both dimensions. We thus define y'_{it} as $y'_{it} = y_{it} * (1 - rc_{it})$. For this measure, larger shares of repeat calls rc_{it} are interpreted as a penalty in performance, since each additional percentage point in the share of repeat calls (lower quality) relates to a lower composite productivity y'_{it} .

3 The role of peers in estimating tenure-performance profiles

Table 3.7: The effect of tenure on quality

Outcome	$\log(y_{it}^q)$ (1)	$\log(y'_{it})$ (2)	Grade 1 (3)	Grade 2 (5)	Grade 3 (5)
log(tenure)	-0.0193*** (0.0031)	0.0332*** (0.0020)	0.0046 (0.1473)	-.0659 (0.1420)	0.0236 (0.2035)
Previous employment	-.0325*** (0.0125)	0.0326*** (0.0106)	0.3549 (0.3977)	0.3817 (0.2570)	0.5665 (0.4293)
Working hours	-.0013*** (0.0004)	0.0001 (0.0002)	-.0081 (0.0127)	-.0020 (0.0107)	-.0021 (0.0135)
Gender (1=male)	0.0027 (0.0084)	0.0074 (0.0067)	0.2920 (0.1874)	0.2629 (0.1865)	0.2355 (0.2642)
Week fixed-effects	Yes	Yes	Yes	Yes	Yes
Observations	7133	7133	736	745	674
Number of agents	330	330	103	103	96
R ²	0.2765	0.2882	0.0507	0.1068	0.0507

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All standard errors are clustered at the agent level. Call quality y_{it}^q is measured as $1/(\text{share of repeat calls} * 10)$. The composite productivity measure y'_{it} is defined as $y'_{it} = aht_{it} * (1 - \text{share of repeat calls})$. Grades are given on ‘knowledge of agent’ (*Grade 1*), ‘agent understood question’ (*Grade 2*), and ‘solution of the problem’ (*Grade 3*). Grades are given on a scale of 1 (very bad) to 10 (very good).

The results, shown in Column (2) of Table 3.7, indicate that there is a positive effect of tenure on the overall measure. The estimated coefficient of 0.0332 is significantly different from zero. A 10% increase in tenure thus leads to an increase in overall productivity of 4.5% of a standard deviation. This suggests that the positive quantity effect outweighs the negative quality effect.

Second, we use information from a customer satisfaction survey as an alternative measure of call quality. This survey is conducted monthly in order to monitor the quality provided by the departments on an aggregate level. Among all incoming calls, customers are selected and called for a short survey. Though there are only about 200 calls evaluated in this way each month, the sample consists of a randomly chosen subset of calls. The survey contains three outcomes related to the agent’s quality. Customers were asked to grade the ‘knowledge of agent’ (Grade 1), whether the ‘agent understood question’ (Grade 2), and whether the agent had a ‘solution to the problem’ (Grade 3) on a scale from 1 (very bad) to 10 (very good).

Columns (3) through (5) of Table 3.7 show that, controlling for an agent’s previous employment, the number of working hours, the gender, and week fixed-effects, there is no significant relation between agent’s tenure and service quality. Neither the knowledge of the agent, nor whether the agent understood the question or had a solution to the customer’s problem appears to be related to tenure.

Overall, these results suggest that the more efficient calls of agents with longer tenure come at the costs of a lower call quality as proxied by repeat calls. Overall, however, our results suggest that performance increases with tenure.

3.4.4 Working hours and the shape of tenure-performance profiles

In the call centre we analyse, there is substantial within- and between-worker heterogeneity in working hours. If learning on-the-job is related to the actual number of hours worked, this heterogeneity may lead to differences in the shape of estimated tenure-performance profiles, depending on the number of working hours.

In order to assess this possible bias, we generate two alternative measures of tenure: tenure measured in working weeks (as opposed to calendar weeks) and tenure measured in full-time equivalents (fte). Tenure in working weeks is defined as the number of weeks an agent worked in the past apart from holiday weeks and other non-working weeks. The second measure, tenure in fte, is defined as the cumulated number of hours worked, divided by 38. It thus measures the number of fte weeks the agent has been working for the call centre. In order to make the variables comparable to the original tenure variable, both variables are rescaled to years. As it is important to have the full history of the actual working hours of agents for this analysis, we only use new hires.

Table 3.8 shows the estimation results of regressions of log-performance on the three different tenure definitions. Column (1) shows the estimation results of our baseline with this smaller sample. The baseline estimate using the standard definition of tenure is 0.0444, compared to 0.0442 using the full sample (Column (5) of Table 3.2). The performance effect of tenure in working weeks in Column (2) is slightly higher (1.58%) than in the analysis with the standard measure of tenure. Since the difference between the two estimates refers to non-working weeks, the result suggests that the tenure-performance profile is actually slightly steeper, than if the non-working weeks are neglected. When using the third definition of tenure, tenure in fte weeks, the effect of tenure appears to be significantly lower than the effect of tenure measured in working weeks. This suggests that agents with more working hours have a tenure-performance profile that is less steep than the tenure-performance profile of workers who work less hours. This result may be explained by sorting high-ability individuals (e.g., students in higher education who work for a shorter time in the call centre) into contracts with shorter working hours.

3 The role of peers in estimating tenure-performance profiles

Table 3.8: Determinants of performance using different tenure definitions

	(1)	(2)	(3)
log(tenure in calendar weeks)	0.0444*** (0.0026)		
log(tenure in working weeks)		0.0459*** (0.0026)	
log(tenure in fte weeks)			0.0409*** (0.0025)
Previous employment	0.0573* (0.0315)	0.0496 (0.0350)	0.0491 (0.0340)
Working hours	0.0002 (0.0004)	0.0001 (0.0004)	-.0008** (0.0004)
Gender (1=male)	0.0017 (0.0085)	0.0013 (0.0084)	0.0016 (0.0084)
Constant	0.3677*** (0.0125)	0.3745*** (0.0124)	0.4140*** (0.0131)
Week fixed-effects	Yes	Yes	Yes
Observations	6451	6451	6451
Number of agents	303	303	303
R ²	0.2953	0.3043	0.2933

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: y_{it} . All standard errors are clustered at the agent level. All regressions are estimated using observations with a maximum of one year of tenure. ‘Tenure in working weeks’ is defined as the number of working weeks an agent worked. ‘Tenure in fte weeks’ is defined as the number of fte weeks an agent worked. All regressions include new hires only which are defined as agents who can be observed from their first working week.

3.5 Conclusions

In this chapter, we analysed tenure-performance profiles of call centre agents, using unique panel data of a multi-national telecommunication firm's call centre that contain detailed performance information on the individual level. The results show that call agents have a very steep tenure-performance profile in the beginning of their job in the call centre. This increase, however, flattens considerably after about 6 to 8 months. The tenure-performance profile can very well be described by a logarithmic function. An increase of tenure by 10% leads to an increase of 5.5% of one standard deviation in performance.

We interpret this increase in performance as learning due to accumulation of human capital. For call agents working for this call centre, the estimated shape of the tenure-performance profile makes sense. The agents' task itself—answering customer calls—is repetitive. The highest returns to learning should therefore occur in the beginning of the employment relationship. In our data, agents perform 66.5% better after 8 months which is comparable to related findings from other sectors. After this initial period, the marginal return to tenure decreases. The rather fast learning of agents may be due to the relatively low complexity of the job. In other, more complex jobs, tenure-performance profiles may be less steep. The initially lower pay-off of learning on-the-job can be recouped over longer job spells. These results imply that the firm's investment in learning on-the-job is substantial. The agents' increase in performance translates to a learning investment by the firm of about 162 hours. Over the same period, the firm's investments in formal training are much lower and accumulate to an average of 84 hours.

In addition, we show that workers significantly benefit from working with more experienced peers. We find that an increase in the average team tenure of one standard deviation relates to an increase in performance of 14% of one standard deviation. This shows that it can be beneficial for firms to place new workers in teams with more experienced agents and thereby facilitating knowledge spill-overs between workers even if the workers are not involved in team work.

4 The effects of training on own and co-worker productivity

4.1 Introduction

Estimating the returns of work-related training on productivity is a challenge for two reasons.¹ First, the non-random selection of workers into training makes it difficult to identify the causal effect of training participation on individual productivity from other unobservable factors that drive participation in training as well as productivity. If this selection is not appropriately accounted for, the causal impact of training programmes on productivity can be overestimated (Bassanini *et al.*, 2007). Second, it is difficult to find appropriate proxy variables for productivity. While wages are often used to estimate returns to training participation, the returns of employer-provided training on productivity should be higher than those on wages (Dearden *et al.*, 2006); however, direct measures of productivity at the individual level are scarce.

Our contribution to the literature is twofold: We are the first to provide experimental evidence on (1) the causal effects of investments in training on productivity using individual performance information and (2) the externalities of training on co-worker productivity. A randomised field experiment enables us to measure the causal effects of investments in training on worker productivity. The field experiment was carried out in the call centre of a multi-national telephone company in the Netherlands. We have detailed information on the contents, length, and purpose of the training programme, as well as unique panel data with administrative information on individual performance, to estimate the returns to training. Agents were randomly assigned to treatment and control groups. This exogenous variation in training participation is used to estimate the returns that are causally related to training and not to unobservable factors affecting both training participation and productivity. We find that agents perform 10% better after participating in the training programme than non-treated agents. We show that this effect is causal and not the result of employee selection into and out of training. Furthermore, we find that the performance increase is not due to a lower quality of work.

Workers belonging to the treatment group were trained over successive weeks. Moreover, due to the restricted capacity of the training centre the teams had to be split into two training groups. This time-varying treatment of the agents in a team allows for the further random assignment of agents into a first and a second

¹This chapter is joint work with Andries De Grip (De Grip and Sauermann, 2012). We would like to thank Iwan Barankay, Eric Bonsang, Lex Borghans, Thomas Dohmen, Dan Hamermesh, Ben Kriechel, Olivier Marie, Eric Maurin, Raymond Montizaan and three anonymous referees for helpful comments. This chapter was presented at Maastricht University, Education in Adulthood and the Labour Market at IAB, IWAAE 2010, EEA 2010, RWI, Advances with Field Experiments (Wharton School), RES 2011, SOLE 2011, IZA Summer School 2011, and ZEW.

training group. This enables us to identify possible externalities from training on untreated peers within a team. We find that a 10 percentage point increase in the share of treated agents leads to a performance increase of 0.51% for their untreated teammates.

From a policy point of view, an unbiased estimation of the impact of training on productivity is important for assessing the role of further training in the development of human capital. Previous studies have found rather mixed results, mostly depending on the method applied, as well as the measure of productivity used. We therefore categorise the research into three main strands. A large number of studies focus on the wage effects of work-related training and are extensively surveyed in Bartel (2000), Dearden *et al.* (2006), and Bassanini *et al.* (2007). After correcting for selectivity by applying fixed-effects estimation or instrumental variables regressions, these studies find returns of training ranging from zero (Goux and Maurin, 2000) to 10% (Parent, 2003). The disadvantage of using wages to measure the effects of training, however, is that wage increases after training only reflect the returns to the worker. The effect on productivity is thus likely to be underestimated under rent sharing.

The second strand of research uses large-scale surveys across firms, establishments, or industries. As a measure of firm productivity, most of these studies use the value added or sales of firms (Black and Lynch, 2001; Dearden *et al.*, 2006; Zwick, 2006; Konings and Vanormelingen, 2010) or direct measures of productivity within one sector of industry (Holzer *et al.*, 1993). Both Dearden *et al.* (2006) and Konings and Vanormelingen (2010) show that the effects of training on firm productivity are about twice those on worker wages. When controlling for various sources of worker heterogeneity, the latter authors find that the productivity premium for a trained employee is, on average, 17%. Though large-scale surveys can have the advantage of providing a representative sample for one or more sectors of industry, they inherently suffer from unobservable heterogeneity in the type as well as duration of training programmes and firm production processes (Ichniowski and Shaw, 2012). Moreover, it is difficult, if not impossible, to find direct measures of productivity that are comparable across industries.

The third strand of research on the effects of training on productivity focuses on just a single firm or establishment, or comparable establishments within a sector of industry (Bartel, 1995; Krueger and Rouse, 1998; Liu and Batt, 2007). Using direct performance measures and wages, they find mostly positive effects of training.

Overall, previous studies have estimated a range of treatment effects of training participation on productivity. At the same time, there is consensus that appropriate

4 The effects of training on own and co-worker productivity

correction for selection into training matters for estimation. Although the use of field experiments has sharply increased over the last decade, only a few studies in the human capital literature exploit (quasi-) exogenous variations in training participation to estimate the effects of training. Leuven and Oosterbeek (2008) use information on the reasons for non-participation to artificially create treatment and control groups and find no significant effects from participation in training. The authors conclude that the usual methods to correct for selection into training do still overestimate the true effect of training. Similarly, Schwerdt *et al.* (2011) find no significant effects in their study analysing the effect of randomly assigned government-funded training vouchers.

This chapter exploits random assignment to training by means of a field experiment to overcome selectivity in training participation. By randomly assigning agents to treatment and control groups, we can estimate the average treatment effect (ATE) of training participation on individual productivity. We use panel data on individual worker productivity to estimate the effects of the training programme. To measure productivity, we use the key performance indicator used by the call centre to evaluate its call agents, that is, the average time needed to handle inbound customer calls. This performance measure has also been used in other studies in the call centre sector (Liu and Batt, 2007; Breuer *et al.*, 2010). We measure each worker's performance each week before and after the training, which allows us to analyse short-run performance dynamics in the weeks after training. Using individual productivity as an outcome allows us to capture the total effect of training on productivity, while wage information captures only the share an agent receives.²

Despite a growing literature on peer effects in the workplace, the externalities of training have hardly been discussed in the literature on the impact of training on productivity.³ The literature on human capital externalities originates from growth theory and has been applied to human capital spillover effects from workers' level of education (Moretti, 2004). When the returns to training are estimated at the individual (worker) level without taking knowledge spillovers into account, returns will be underestimated. This also holds for a second type of externalities, peer effects due to social pressure (Falk and Ichino, 2006). The identification of externalities,

²A second advantage of using individual productivity instead of annual productivity data is that studies using annual data are more likely to underestimate the causal effects of training. If the yearly data do not contain information on the training period and use average productivity over the previous year, the returns to training are likely to be underestimated, even if selectivity is properly controlled for.

³To the best of our knowledge, the only exception is the study of Dearden *et al.* (2006) who aim to estimate externalities by comparing individual-level wage regressions with industry-level wage regressions that should capture externalities.

however, is empirically difficult. Previous studies used either quasi-random variations in group composition (Mas and Moretti, 2009; Chapter 3 of this thesis) or the random assignment of subjects to groups (Guryan *et al.*, 2009). While many studies have exploited changes in management behaviour induced by firm management (e.g., Shaw and Lazear, 2008; Bandiera *et al.*, 2007), we exploit the structure of the field experiment to estimate externalities. To achieve identification, teams were randomly split into two training groups. This random assignment of agents from the same team allows us to estimate the causal estimates of within-team externalities.

The training we analyse was a one-week course to train call agents in conversation techniques designed to decrease the average handling time of calls while maintaining call quality. Call centres have become a major sector of employment due to strong growth rates since the 1980s, facilitated by the increasing availability of information and communication technology infrastructure. For the US, Batt *et al.* (2005) estimate that call centres employed about 4 million employees, roughly 3% of the total workforce. Although the current trend is to outsource call centres, most call centres are in-house (Batt *et al.*, 2009). Work-related training is an important element of the call centre industry. In general, call agents receive hardly any initial vocational training, whereas the heavy use of information technology in in-house centres requires high investments in work-related training. From this perspective, we expect the returns on training in call centres will be high (Sieben *et al.*, 2009).

The remainder of this chapter is structured as follows: Section 4.2 provides an overview of the firm's workers and describes the experiment in detail. Section 4.3 presents our regression model and the estimation results. Additional evidence and tests on the robustness of our results are discussed in Section 4.4. Finally, Section 4.5 summarises the chapter and presents concluding remarks.

4.2 Context of the experiment

4.2.1 Organisation of the department and worker tasks

The field experiment analysed in this chapter was implemented in an in-house call centre of a multi-national mobile network operator in the Netherlands. The call centre acts as a service centre for current and prospective customers. It has five departments segmented by customer group. To ensure a homogeneous production process, we focus on the largest department, that for private customers with fixed cell phone contracts.

Table 4.1: Selection into the experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Selection into the experiment ($N=179$)		Initial assignment ($N=86$)		Estimation sample ($N=74$)					
Agents:	All	Ass.	Non-ass.	diff (2)-(3)	TG	CG	diff (5)-(6)	TG	CG	diff (8)-(9)
Gender	.2905	.2791	.3011	-.0220	.3513	.2245	.1269	.3824	.2750	.1074
(<i>share of male agents</i>)	(.4553)	(.4512)	(.4612)		(.4840)	(.4216)		(.4933)	(.4522)	
Age	32.595	35.575	29.431	6.143***	34.245	36.578	-2.332	34.070	36.146	-2.076
	(11.325)	(11.062)	(10.792)		(10.017)	(11.791)		(10.095)	(11.640)	
Tenure	2.646	4.312	1.106	3.206***	4.233	4.372	-.139	4.071	3.773	.298
(<i>in years</i>)	(3.610)	(4.019)	(2.310)		(3.796)	(4.218)		(3.810)	(3.984)	
Performance	.3307	.3716	.2928	.0789***	.3715	.3717	-.0002	.3596	.3582	.0013
(y_{it})	(.1108)	(.1017)	(.1058)		(.0884)	(.1116)		(.0779)	(.0893)	
Call quality	.4541	.4676	.4416	-.0260	.4759	.4609	.0150	.4802	.4571	.0230
(y'_{it})	(.1244)	(.1447)	(.1010)		(.1812)	(.1087)		(.1844)	(.1089)	
Composite productivity	.2529	.2873	.2209	.0664***	.2846	.2895	.0049	.2759	.2780	-.0022
(y'_{it})	(.0864)	(.0818)	(.0783)		(.0687)	(.0916)		(.0583)	(.0705)	
Average working hours	18.000	16.593	19.301	-2.708**	15.270	17.592	-2.322	15.706	17.350	-1.644
	(8.455)	(8.058)	(8.646)		(8.875)	(7.317)		(9.104)	(7.277)	
Share peak hours	.5541	.5304	.5773	-.0468*	.5445	.5197	.0248	.5553	.5286	.0266
	(.1766)	(.1946)	(.1547)		(.1972)	(.1939)		(.1926)	(.1970)	
Number of agents	179	86	93		37	49		34	40	

Significance levels: Difference significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; two-sided t test. Standard deviations are shown in parentheses. Descriptive statistics are calculated before agents were assigned to the training (week 50/2008). TG=treatment group; CG=control group.

The main task of call agents in this department is to answer customer phone calls. Customers contact customer service when they have problems, complaints, or questions. All agents take part in a training course when entering the department that enables them to handle basic types of calls. Throughout their careers, agents receive further training. These training programmes mainly focus on information in promotional campaigns, communication, and information technology skills, as well as on handling more complex calls.

In the first period of our sample, 179 individuals were working in the department. Column (1) of Table 4.1 shows the descriptive statistics of all agents working in the department at the beginning of the period of observation. Most agents are part-timers. The average number of hours worked per week is 18.0, and only 22.4% of agents work 25 hours or more each week. A total of 29.1% of all agents are men, the average age of the agents is 32.6 years, and the agents have an average of 2.7 years of experience working for the firm.

The agents are organised into 10 teams. In general, all teams provide all services, that is, there is no team specialisation in handling certain types of calls or customers. All teams work on the same floor of the building. Work places are organised into work islands, with up to eight agents of a team sitting next to each other. Each team is led by a team leader who reports to the head of the department. The team leaders' main task is supervising and monitoring the agents of their team. They evaluate their agents regularly based on performance scorecards containing detailed information on key performance indicators. Agent pay is based on a single collective agreement. Agents are paid a base wage and receive an annual wage raise of zero to 8%, depending on an annual performance rating by the team leader.

4.2.2 Training purpose, contents, and organisation

The training programme analysed in this chapter was intended for all agents of the department with some experience on the job. Its aim was to increase the efficiency of agents answering customer calls. Agents who were recently recruited were not selected for this training (yet). The management had decided to organise the training to decrease the average time needed for handling calls, because the call centre had been performing below the targets set.

The training was organised as a week-long programme. Due to capacity constraints, only one group, with a maximum of 10 agents, could be trained at a time. The training took place in an in-house training centre located on a different floor in the same building. The programme consisted of 10 half-day training sessions that

4 The effects of training on own and co-worker productivity

were held on five consecutive days, from Monday to Friday. In half of these sessions, agents were either formally trained by a coach or had group discussions assisted by a coach and their team leader. In these group discussions, agents discussed the skills they lacked, how their skills could be improved, and how agents could help each other on the work floor. In formal sessions, agents were trained in conversation techniques designed to decrease the average time needed for handling customer calls without lowering quality. This included, for instance, the way in which call agents gathered information from customers to resolve calls quickly.

The remaining sessions consisted of learning by doing, by either handling regular customer calls or listening to the calls of other call agents. During these sessions, incoming customer calls were routed to the training centre and agents handled these calls under the supervision of the training coach and their team leader.

4.2.3 The field experiment

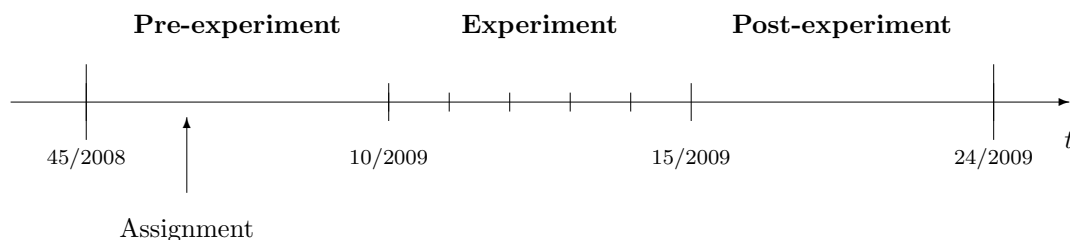
Identification strategy

In the economic literature on estimating the effects of training, the most evident problem is the potential correlation of unobservable factors with both training participation and the outcome variable. This chapter uses exogenous variations in training participation to identify the causal effect of training on individual productivity.

As shown in Figure 4.1, the field experiment consists of three periods. We observe 32 weeks, from week 45/2008 through week 24/2009. At the beginning of the first period (pre-experiment), which lasted 17 weeks, agents were assigned to treatment and control groups. In this period, neither the treatment group agents nor the control group agents were trained. During the second, experiment period, the treatment group was trained consecutively over five weeks. After the experiment period (post-experiment, lasting 10 weeks), agents from the treatment group as well as their untreated peers from the control group worked as usual. Because the agents of the control groups were trained after the post-experiment period (from week 25/2009 onwards), we use data from week 45/2008 through week 24/2009 only. Agents who are part of the control group thus never participate in training throughout our observation period.

The training had been designed for agents with some experience in the job. Out of the 179 individuals working in the department at the beginning of our observation

Figure 4.1: Overview of the field experiment



Note: This figure shows the design of the field experiment. The observations start in week 45/2008 and end in week 24/2009. The assignment to treatment and control groups took place in week 50/2008. Training of the treatment group was held in the experiment period.

period (week 50/2008), 86 were selected for the training.⁴ Management focused the training programme on the more tenured agents for two reasons: First, agents should be experienced to benefit from this type of training and, second, management was aiming to lower the risk of losing its training investments through turnover.

Column (4) of Table 4.1 shows the differences between assigned and non-assigned agents. Assigned agents are an average of six years older, have longer tenure by more than three years, and perform better. While this selection is clearly non-random, it does not violate the assumption that assignment to the treatment group is exogenous. This is because assignment to the treatment and control groups is exogenous conditional on being assigned to the training programme. An advantage of the focus on more experienced agents is that the calls they handle are relatively homogeneous: Agents with longer tenure deal with all types of calls, whereas starting agents do not work on all types of calls yet.

Conditional on being selected for the training programme, 37 of the 86 agents were randomly selected for participation in the treatment group. The remaining 49 agents, who were assigned to the control group, were trained after the post-experiment period, that is, after the end of the observation period. The differences in observable characteristics between the assigned treatment and control groups are relatively small, with none significantly different from zero (Column (7) of Table 4.1). For the estimation, however, agents were excluded if the management reassigned them from the control group to the treatment group, or vice versa, because they could not participate in the planned training week due to vacation or illness. Three

⁴Agents not selected initially were assigned to training weeks after the training of the treatment group. Because these agents could not be randomly assigned to treatment and control groups, we do not consider them to be part of the control group. The effects we find therefore only hold for more experienced agents. As Sieben *et al.* (2009) show, training courses for more experienced agents are highly relevant in the call centre sector.

4 *The effects of training on own and co-worker productivity*

agents from the treatment group and nine agents from the control group are thus excluded from the estimation sample. Column (10) of Table 4.1 shows that there is no statistical difference in observable variables between the treatment and control groups of our estimation sample.⁵ The resulting estimation sample is an unbalanced panel of 74 agents during 32 weeks. In total, we have 1,859 agent-week observations on performance. On average, agents work 28 weeks out of the 32 weeks. The time-variant treatment in our experiment implies that the treatment and control groups differ during each specific week, that is, agents belong to the control group until they are treated. For this reason, teams were not only randomly assigned to the treatment and control groups, but also randomly assigned to the various weeks during the treatment period.

Identification of externalities

Although agents carry out individual tasks, social interactions can arise, influencing the treatment effect.⁶ Agents belonging to the same team sit next to each other. Furthermore, they can communicate with their co-workers during team meetings and breaks. In a survey held among the agents, 80% stated that they exchange work-related information either during official team meetings or at their workplace. Therefore, the likelihood of externalities is higher within teams than across teams. We do not preclude externalities between teams; however, given the physical distance on the work floor, these externalities should be less than externalities within teams.

Our field experiment allows us to identify these externalities of the training on untreated peers. For this purpose, we exploit the fact that eight of the 10 teams had to be split into separate training groups to match the size restriction of training groups in the firm's training centre, and the agents in these teams were randomly assigned to the two training groups.

Divided teams can thus be distinguished in the first and second training groups. In between the training weeks of the two training groups in which a team had been

⁵In addition, all tests were conducted by using the non-parametric Wilcoxon-rank test, which is suited for comparing discrete measures between small groups (Wilcoxon, 1945). The results do not differ from those of the two-sided t -test.

⁶The production technology, however, is not necessarily without externalities. Agents may follow a strategy of providing low quality to score well on the main performance criteria (average handling time). If low-quality calls lead to a higher rate of customers calling repeatedly, other agents can be affected if they have to take more care of those customers. In practice, however, agents cannot easily shirk by providing low quality with a low average handling time because the quality of agents is regularly monitored by their team leaders, who can listen in on the calls. Moreover, if there was any shirking by providing low quality, these agents should be similarly distributed to the treatment and control groups.

Table 4.2: Descriptive statistics of teams' first and second training groups (treatment group only)

	(1)	(2)	(3)
Agents:	First group	Second group	diff (2)-(3)
Gender (<i>share male agents</i>)	.3846 (.5064)	.3810 (.4976)	.0037
Age	35.659 (10.759)	33.085 (9.799)	2.575
Tenure (<i>in years</i>)	3.707 (3.978)	4.297 (3.784)	-.590
Performance (y_{it})	.3687 (.0685)	.3539 (.0843)	.0148
Call quality (y_{it}^q)	.4526 (.0798)	.4972 (.2269)	-.0446
Composite productivity (y'_{it})	.2844 (.0538)	.2706 (.0616)	.0139
Average working hours	18.539 (10.674)	13.952 (7.743)	4.586
Share peak hours	.5181 (.2242)	.5783 (.1720)	-.0602
Number of agents	13	21	

Significance levels: Difference significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; two-sided t test. Standard deviations are shown in parentheses. Descriptive statistics are calculated before agents were assigned to the training (week 50/2008).

split, one or two groups of a different team were trained. During these weeks, the team consisted of exogenously chosen agents who were treated and randomly chosen agents who were not yet treated. This further randomisation enables us to distinguish between three groups of agents: (1) those who were trained first (first training group), (2) their teammates who were trained later (second training group), and (3) the agents in the teams in the original control group who will not be trained in the period we analyse. We exploit the fact that the agents of one team were randomly assigned to training groups to identify within-team externalities. Table 4.2 shows the means of observable characteristics for treatment group agents in a team's first training group ($N=13$) versus the means for agents who are in a team's second training group ($N=21$).⁷

All agents were informed about the training at the moment of their initial assignment. Throughout the experiment, information about the randomisation of the

⁷Both the t -test as well as the Wilcoxon-rank test do not show significant differences between the groups. It should be noted, however, that the sample size for this comparison is rather small.

4 The effects of training on own and co-worker productivity

order of the teams was given only to the head of the training department and the department manager. No information was given to the team leaders or the agents about the randomisation or the evaluation of the training programme. There was no need to do this, since all employees know that the size of the training centre is restricted. For this reason the training department always makes a training schedule in collaboration with the department managers when they start a new training program. The training centre is occupied most of the time. During the post-experiment period, agents from other departments of the call centre were trained in the training centre.

4.2.4 Measuring productivity

We measure agent productivity by means of the call centre's key performance indicator: the average handling time, defined as the average time an agent needs to handle a customer call (Liu and Batt, 2007; Breuer *et al.*, 2010). The handling time includes the time needed to talk to the customer, as well as the time needed to log the call in the customer database. Information on the average handling time of each individual agent i is available for each working week t . This performance measure does not contain information on the quality of calls. Section 4.4.3 introduces additional indicators of call quality and discusses the quality-quantity relation.

The department's aim is to improve performance by decreasing average handling time. Throughout this chapter, we therefore use the inverse of the average handling time (ah_{it}) multiplied by 100: $y_{it} = \frac{1}{ah_{it}} \cdot 100$. Since lower values of ah_{it} are interpreted as higher performance, this transformation allows us to interpret improvements in y_{it} as increases in individual productivity. The average performance of all agents assigned to training is 0.3307, which relates to 5.6 minutes for an average call (see Column (1) of Table 4.1).

There is substantial heterogeneity in individual performance within and between agents. This suggests that not only individual-specific characteristics, but also other, department-specific effects such as technical problems affect the individual productivity of all agents working in the department. Compared to individual heterogeneity in productivity, however, variation over time is less important: While period fixed effects explain only about 12% of the overall variation in individual performance, worker fixed effects alone explain 43% of the overall variation in performance.

It is essential that our measure of productivity is comparable within and between agents. Calls are randomly assigned to agents. Agents have no direct influence on the types of calls they receive or the types of customers put through to them, and,

therefore have the same probability of exceptionally long or short calls. Before talking to agents, however, customers must first state the purpose of their call. Based on this information, calls are routed to agents who have sufficient knowledge to resolve the call. The assignments of agents to types of calls can be changed at any time by management. Agents are often reassigned *ad hoc* if the structure of customer calls changes.

This assignment of calls to agents does not violate the assumption that our performance measure is comparable across time and agents. First, because agents are exogenously assigned to the treatment and control groups, the skill distribution of agents and thus the types of calls should, on average, be the same in the treatment and control groups. Second, we compare an agent's performance before and after the training intervention. Because the training did not focus on resolving different types of customer requests, calls assigned to agents after the treatment do not systematically differ from those assigned prior to the training.

4.3 Empirical analysis

4.3.1 Causal effect of training on individual performance

We observe an agent i 's performance y_i and training participation d_i . The observed outcome can thus be written as

$$y_i = y_i(d_i) = d_i \cdot y_i(1) + (1 - d_i) \cdot y_i(0) \quad (4.1)$$

where $y_i(1)$ and $y_i(0)$ denote performance in the treated and untreated states, respectively. The randomised assignment of agents to the treatment and control groups ensures the independence of treatment status and potential outcomes $E[y_i(j)|d_i = 0] = E[y_i(j)|d_i = 1]$ for $j = 0, 1$. The ATE τ (Rosenbaum and Rubin, 1983) is thus identified by $\tau = E[y_i(1) - y_i(0)]$. The ATE can be estimated by performing a linear regression of individual i 's performance y_{it} in week t on a treatment dummy d_{it} , which is defined as being one in each after-training period, and zero otherwise:

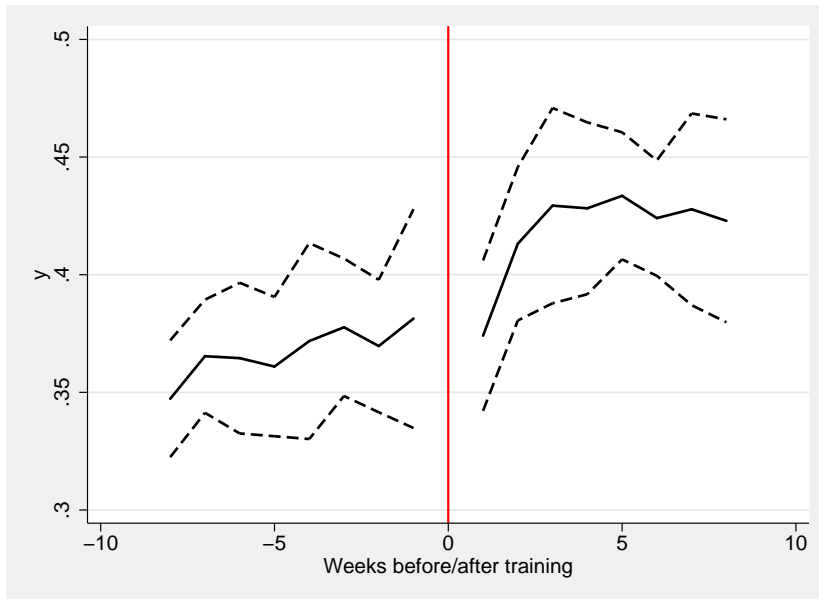
$$y_{it} = \alpha_i + \tau d_{it} + \beta_1 t_t + \beta_2 X_{it} + \beta_3 X_t + u_{it} \quad (4.2)$$

where α_i are individual fixed effects to account for individual heterogeneity remaining despite the experimental design, and X_{it} are covariates that are assumed to be independent from the treatment status d_{it} , such as working hours in week t and the

4 The effects of training on own and co-worker productivity

type of shift the agent is working.⁸ The variable t_t is a linear time trend that controls for trends in aggregate performance affecting all agents, and u_{it} is an idiosyncratic error term. To control for aggregate effects on performance (X_t), we include the overall number of calls divided by the total number of full-time equivalent agents. Throughout this analysis, standard errors are clustered at the agent level.⁹

Figure 4.2: Average performance of the treatment group before and after the training



Note: This figure shows the average performance of agents in the treatment group in the weeks before and after the training week (solid line). The dashed lines show the 95% confidence interval. The training week is denoted as week 0.

4.3.2 Baseline results

We provide first evidence of the treatment effect in Figure 4.2, which shows the average performance of agents in the treatment group. The treatment week is denoted week 0, with positive (negative) values of the x axis showing the t th calendar week after (before) the training. Agent performance appears to be, on average, higher in the weeks after the training than before the training. When the performance of un-

⁸We do not use agent tenure as a covariate in our regressions for two reasons. First, because the agents in our sample are all relatively experienced, there is not sufficient variation in individual tenure to control for tenure *and* for common time trends. Second, previous research shows that experience effects matter only for new agents (Chapter 3 of this thesis).

⁹In panels increasing in T with fixed N , the appropriate assumptions about standard errors can be crucial to the significance of the results (Bertrand *et al.*, 2004). The results here were re-estimated, allowing for serial correlation in the standard errors. This does not change the size or significance of the results.

Table 4.3: Average performance post-treatment period

	(1)	(2)	(3)
	Treatment group	Control group	Difference TG-CG
Average performance (predicted)	.4148 (.0143)	.4027 (.0157)	.0122***
Number of agents	28	35	
Number of observations	296	409	

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; two-sided t test. Standard deviations in parentheses. Predicted performance is generated from a regression of performance on working hours of agent i in week t plus a linear time trend.

treated agents is not controlled for, treated agents perform significantly better after the training. Given the random assignment of agents to the treatment and control groups, this can also be shown by comparing the mean performance of agents in the treatment group with that of agents in the control group. Table 4.3 shows that agents in the treatment group are significantly more productive after the training, while there are no significant pre-treatment differences between the two groups.

Table 4.4 shows the results when estimating Equation (4.2). The treatment dummy is defined as being one in all weeks after an agent has been trained, and zero otherwise. While agents from the control group thus always have a treatment dummy equal to zero, the share of treated agents increases in time with the growing number of groups that have been trained. Column (1) of Table 4.4 shows that agent performance after participation in the training is 11.3% higher than before the training, controlling for untreated agents' performance. When, in addition, individual heterogeneity is controlled for by including worker fixed effects, the effect increases slightly to 12.4% (Column 2). Figure 4.2, however, suggests that aggregate trends matter. When controlling for a linear time trend, the effect decreases by about 3.6% to an estimated effect of 8.8% (Column 3 of Table 4.4).¹⁰ The time trend itself is significant, suggesting that there is an overall positive trend in performance during the sample period.

The estimates in Column (3) of Table 4.4 show that the number of working hours do not have a significant impact on the performance y_{it} . In order to control for possible differences in the nature and amount of calls during peak hours, we use a variable that contains the share of peak hours in the total number of hours agent i worked in week t . The estimates show that a 10 percentage-point increase in the share of peak hours leads to a decrease in performance of 3.4%. This suggests that

¹⁰To allow for more flexibility in time trends common to all agents, we also estimated Equation (4.2) including linear and squared time trends. The estimated treatment effect $\hat{\tau}$ then decreases slightly. The difference between the two estimates is not significantly different from zero.

4 The effects of training on own and co-worker productivity

Table 4.4: Direct treatment effect

	(1)	(2)	(3)
Treatment dummy	.1127*** (.0295)	.1244*** (.0152)	.0882*** (.0206)
Working hours	.0017 (.0018)	-.0012 (.0010)	-.0015 (.0010)
Share peak hours	-.2564*** (.0930)	-.3486*** (.0739)	-.3416*** (.0749)
Calls per FTE	.0002*** (.0001)	.0002** (.0001)	.0001 (.0001)
Time trend			.0023** (.0009)
Constant	-1.0008*** (.0782)	-.8770*** (.0582)	-1.0000*** (.0888)
Individual fixed effects	No	Yes	Yes
Observations	1859	1859	1859
Number of agents	74	74	74
R ²	.0689	.6113	.6165

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All standard errors are clustered at the agent level.

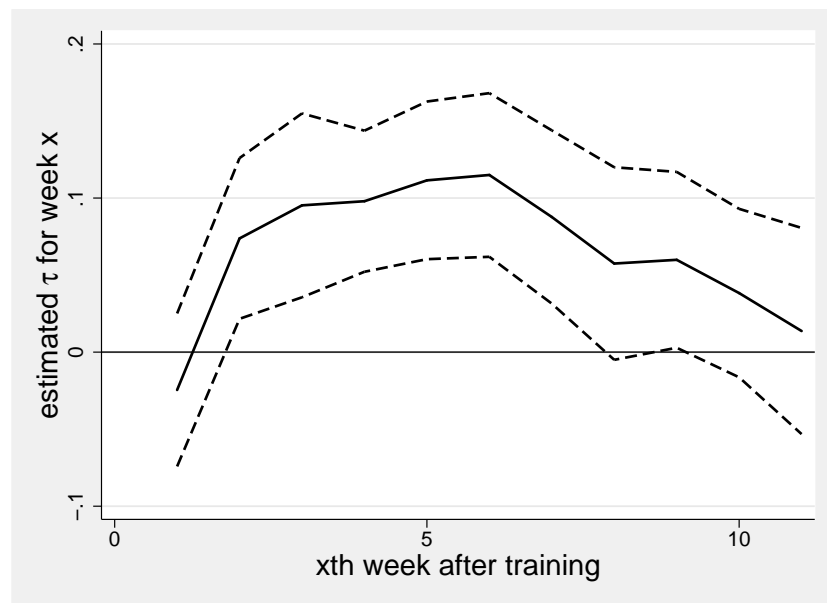
the calls during peak hours are more time consuming than the calls at other times of the day.¹¹ In order to control for overall pressure in week t , we furthermore control for the effect of the overall number of incoming calls in week t , divided by the number of full-time equivalent agents. This variable, however, does not significantly affect the performance of agents.

The preceding analysis makes the standard assumption that participation in training leads to a persistent shift in performance by including a dummy for participation in training. Given the weekly performance information at hand, we can exploit dynamic patterns of the treatment effect. Figure 4.3 shows the estimated shape of the treatment effect in the weeks after the training. Panel (a) of Figure 4.3 is based on an estimation of Equation (4.2) where the treatment variable d_{it} has been replaced by a set of dummies for each post-training week ($\sum_{t'=1}^T d_{it'}$), where $t' = 1$ denotes the first week after the training. The untreated state $d_{it} = 0$ serves as the reference. It shows that the increase in performance in the first week after the training is not significantly different from zero. In the weeks thereafter, agent performance is significantly higher than before the training. Soon after, however, the estimated treatment effect

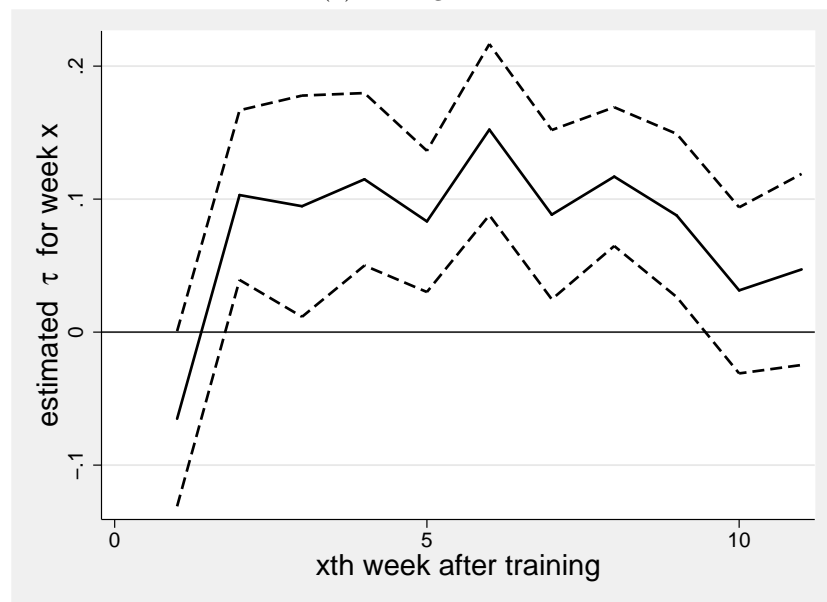
¹¹We also estimated regressions with similar variables to control for the timing of working hours, such as dummies for different working hours. These estimates show very similar results.

gradually decreases and becomes insignificant from about the eighth week after the training. This decrease suggests that the estimate $\hat{\tau}$ can only be interpreted as an average effect over the whole post-treatment period.

Figure 4.3: Development of the treatment effect on performance over time



(a) Average effect



(b) First training groups only

Note: This figure shows the estimated treatment effect on performance for each week after the training, controlling for a linear time trend. Panel (a) shows the estimated treatment effect from a regression including all agents. Panel (b) shows the treatment effect for first groups only. Week 0 denotes the training week. The dashed lines show the 95% confidence interval estimated from a regression of y_{it} on after-training week dummies.

4 The effects of training on own and co-worker productivity

While this result seems to suggest that the treatment effect declines over time, there may be other mechanisms at work. From a human capital perspective, it is less reasonable that human capital acquired during the training depreciates within a short number of weeks. Moreover, the declining treatment effect may be due to a decline in the agents' motivation. From this perspective, however, it is difficult to understand that workers are least productive in the first week after the training. When the treatment effect would be driven by motivation, there should be a high positive performance effect in the first week. It is more likely that the decreasing treatment effect is driven by training externalities. If (yet) untreated agents are affected by the training of their peers, either due to knowledge spillover or peer pressure, their performance will increase when their co-workers have been treated. In this case, the treatment effect measured should be highest in the weeks shortly after the training and then constantly diminish, because agents from the control group are not actually fully untreated but indirectly affected by the training as well. Panel (b) of Figure 4.3 takes account of possible training externalities by showing the estimated shape of the treatment effects in the weeks after the training for the agents in the first training groups only, with the agents who will not be trained throughout the observation period as the control group. This figure shows a similar pattern as Panel (a). The treatment effect, however, remains significant for a longer period. Both the treatment and control groups in this setting are not affected by within-team externalities suggesting that Panel (a) exhibits contamination through externalities leading to a decreasing treatment effect over time.

4.3.3 The role of social interaction

The preceding subsection estimates the ATE from participation in the training programme in the way that is standard in the training literature. An unbiased estimate $\hat{\tau}$, however, requires the stable unit treatment value assumption, hereafter SUTVA (Angrist *et al.* 1996). The most important implication of the SUTVA is that there are no externalities from treated workers on untreated workers; that is, the control group's performance ($y(0)$) is not affected by the treatment. In settings where individuals potentially interact, the SUTVA is violated. In the presence of externalities on untreated agents, the observed outcome changes to

$$y_i = y_i(d_i) = d_i \cdot y_i(1) + (1 - d_i) \cdot (1 + d_i^* \tau_s) \cdot y_i(0) \quad (4.3)$$

where d_i , $y_i(0)$, and $y_i(1)$ are defined as in Equation (4.1). In addition, we allow the observed performance in the untreated state to be affected by the treatment of workers in the same team. The indicator d_i^* is defined as being one if a worker is untreated but exposed to treated workers from the same team. If an agent is untreated ($d_i = 0$) and working in a team in which all agents are untreated ($d_i^* = 0$), the potential outcome from Equation (4.3) simplifies to that from Equation (4.1). The same holds for treated agents who participated in the training ($d_i = 1$ and $d_i^* = 0$). The indirect treatment effect τ_s scales the size of the externality. If we assume that externalities are non-negative and affect untrained agents only, Equation (4.3) implies that, unless $\tau_s = 0$, the estimated treatment effect ($\hat{\tau}$) is underestimated in Equation (4.2). If direct and indirect treatment effects are stable, the true treatment effect is underestimated by τ_s .

In practice, the identification of social effects is cumbersome due to endogeneity (Manski, 1993). In our setting, however, we can exploit the fact that team agents were randomly assigned to two separate training groups. In between the training of two groups belonging to the same team, treated and untreated agents of the same team worked together for one or two weeks. During these weeks, a randomly selected group of the team that was treated worked with the remainder of the team that was untreated. Untreated agents of these teams were thus exposed to treated peers. Assuming that there are no peer effects between teams, the random assignment enables us to identify within-team externalities. We can do this in both a direct and an indirect way. The first column of Table 4.5 shows the results of a direct test of externalities. Here, we test whether the share of treated peers affects the performance of agents who were not yet treated. We therefore use precise information about an individual agent's shift to calculate the share of treated agents among all the agents with whom the agent worked during that shift. The share of treated peers is calculated for peers from the same team only. The results show that a 10% increase in the share of treated peers leads to a performance increase of 0.51%. This effect, however, is only statistically significant at the 10% level of significance.¹²

The second training groups having already increased their productivity through the externality effect of the first group being trained, we can expect a smaller additional effect when the second training groups finally receive the training themselves. We

¹²The regression presented in Column (1) of Table 4.5 does not contain the variable *share peak hours*. This is because both variables *share of peers treated* and *share peak hours* are based on the same hourly information. This causes a high between-agent within week correlation of the two variables and makes it impossible to identify the effect of *share peak hours* in this regression. This is supported by the fact that agents have only limited possibilities to choose shifts, and that the assignment to the first and second training group is random (Table 4.2).

4 The effects of training on own and co-worker productivity

Table 4.5: Estimation of externalities

	(1)	(2)	(3)
Share of peers treated	.0513* (.0275)		
Treatment dummy		.0992*** (.0213)	.0675** (.0287)
Constant	-1.1772*** (.0702)	-1.0548*** (.1042)	-.9924*** (.1018)
Individual fixed effects	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
Observations	1544	1352	1510
Number of agents	74	53	61
Number of agents (treatment group)	34	13	21
Number of agents (control group)	40	40	40
R ²	.5612	.6173	.5894

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All standard errors are clustered at the agent level. *Share of peers treated* is calculated as the average share using information on the actual working hours in a specific week. The sample used for Column (1) is restricted to agents in an untreated state only (treatment dummy=0). Control variables include *working hours*, *share peak hours*, a linear time trend, and *calls per FTE*. The variable *share peak hours* is not used in Column (1).

therefore test whether $\hat{\tau}$ is lower for agents who were exposed to treated agents before their own training. Column (2) of Table 4.5 shows the estimated treatment effect for agents who comprise the first group of their team to be treated, compared to that for the control group of agents who were in teams in which none of the agents were treated throughout the observation period. The estimated effect is slightly higher (9.9%) than when estimating Equation (4.2) for all agents (8.8%; see Column (3) of Table 4.4). This suggests that the causal effect from training is underestimated because some agents were exposed to treated agents in their teams. Accordingly, the same estimation for second-group agents results in a lower estimate. Column (3) of Table 4.5 shows that their treatment effect is 6.8%. The difference between the two estimates, however, is not significantly different at the 5% significance level. Despite this, the results shown in Table 4.5 strongly suggest the existence of peer effects from treated agents on untreated peers in their team.

4.3.4 Returns to training

Standard theory predicts that firms invest in human capital if the expected returns from training investments exceed costs. Personnel information with specific information on the length and costs of training, as well as information on performance,

allows to estimate the returns to training (Liu and Batt, 2007). For the training analysed in this chapter, the costs of training per agent are defined as

$$C = \frac{(n_g \cdot w_a + w_l + w_c) \cdot h}{n_g} \quad (4.4)$$

where n_g is the size of the training group; w_a , w_l , and w_c are the hourly wages for the agents, team leader, and training coach, respectively; and h is the length of the training in hours.¹³ Because our performance measure is the time an agent needs, on average, to handle a call, and there are no other tasks in which the agent is involved, benefits are defined as the percentage increase in performance times the weekly workload and the agent's wage. The benefit B_d can be calculated for each week after the training d :

$$B_d = (\hat{\tau}_d \cdot aht_{id} \cdot n_{c,id}) \cdot w_a \quad (4.5)$$

where $n_{c,id}$ is the number of calls an agent completes in after-training week d and $\hat{\tau}_d$ is the estimated treatment effect for this week as shown in Panel (b) of Figure 4.3. The term B_d is thus the decrease in the time a trained agent needs to handle a certain workload compared to an untrained agent in week d , multiplied by his wage.

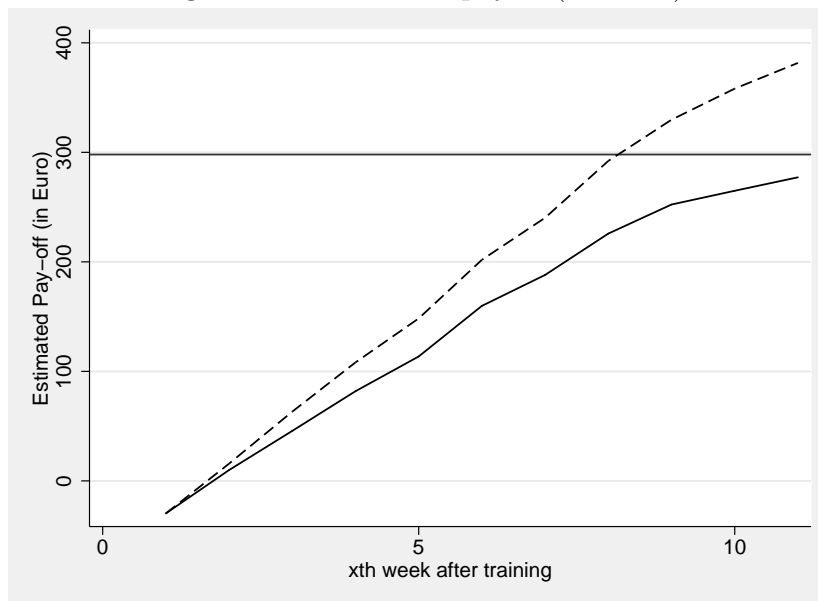
Given the decreasing treatment effect over time, we use the actual average performance and actual number of calls made to calculate the returns to training for the firm. The firm will *de facto* retain the full benefits of the performance increase because agents' wages are almost entirely related to their years of tenure. Figure 4.4 shows the estimated curve of cumulated returns $\sum_{d=1}^D B_d$ until week D . The solid line shows that, for first-group agents, the cumulated benefits do not exceed the costs of the training in our sample period. When considering within-team externalities, however, the training reaches its break-even point in the eighth week after the training.¹⁴ Figure 4.4 shows that in week 11 after the training, the overall return to the training is 37.7% higher when including externalities in the calculation of the benefits.¹⁵

¹³The wage costs of the team leader and training coach who supervise the training are included as direct training costs.

¹⁴Because half of the training consisted of actual customer calls, we calculated the costs of the training using 50% of the agents' wage costs. Considering the full wage costs of agents roughly doubles the training's costs.

¹⁵Externalities were calculated based on the estimates of the treatment effect for the first- and the second-group agents ($\hat{\tau}_1$ and $\hat{\tau}_2$) from Columns (2) and (3) of Table 4.5. We use the difference between the treatment effects as a measure of externalities. The externality effect on second-group agents is thus calculated as $(\hat{\tau}_2 - \hat{\tau}_1) \cdot aht_{id} \cdot n_{c,id} \cdot w_a$.

Figure 4.4: Estimated payoff (in euros)



Note: Cumulative payoff for first-group agents (solid line) and first-group agents including externalities on teammates (dashed line). The horizontal axis shows the individual training costs based on an agent’s wages and the trainer’s and the team leader’s wage costs (see Section 3.4).

4.4 Additional evidence

4.4.1 First-week effects

One finding of the preceding analysis of the dynamics of the treatment effect is that agents perform better only starting from the second week after the treatment (see Figure 4.3). We can rule out the argument that this is due to selectivity, since over 90% of the agents who participated in the training also worked the following week. Another, yet untestable, explanation for the low effect is that the low performance in the first week after the training is due to passing knowledge on to teammates.

A testable explanation, however, is fatigue. Most of the agents work part-time, with an average of about 20 hours per week. The training programme is one full-time week, that is, 38 hours. When using additional data on previous training programmes, we find that after rather short training programmes (less than 15 hours of training in one week), the difference between pre- and post-training is about the same, irrespective of the call agent’s usual working hours (see Column (1) of Table 4.6). This is different for long training programmes where agents with relatively low numbers of working hours perform worse in the week after the program than before. In contrast, agents with long working hours do not exhibit this decrease in performance in the first week following the training. This result suggests that fatigue

Table 4.6: Pre-/post-differences in performance for other training programmes

	(1)	(2)
Length of courses	short	long
Agents with short working hours	-.037 (.080)	-.047*** (.012)
Number of observations	42	526
Agents with long working hours	-.006 (.012)	-.040 (.034)
Number of observations	742	64

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The numbers show the differences in y_{it} between the first week after the training and the week before the training. The calculations are based on a larger data set from the same call centre. Short courses are defined as courses with less than 15 hours per week. Long courses are defined as courses with more than 25 hours per week. Short and long working hours are relative to the course length. Standard errors are in parentheses.

is an explanation for the slight decrease in performance in the first week after the training.

4.4.2 Personnel turnover and training effects

As in other call centres, the call centre analysed in this chapter experiences a high agent turnover. In total, we assigned 86 agents to the treatment and control groups, of which 12 were reassigned by the management or dropped out (see Section 4.2.3). Table 4.1 shows that the treatment and control groups do not differ with respect to observable characteristics such as age, gender, and tenure. This holds for the initial assignment, as well as for agents who eventually participated in the training.

Ordinary least squares estimation of Equation (4.2) results in an unbiased estimator of τ only if the probability of belonging to either the treatment or control group does not depend on unobservable characteristics that are correlated with both the likelihood to participate in the training programme and our performance measure, average handling time. This type of selectivity occurs when agents with these characteristics are more likely to drop out of the treatment group than from out of the control group.

To test this hypothesis, we limit the estimation sample to agents that eventually participated in the training programme and stayed in the department throughout the observation period. This decreases the sample size from 74 agents in our main estimation sample to 55 agents. Of the 34 agents of the treatment group, 26 (77%) were trained and did not leave the call centre. For the control group, this limits the number of agents from 40 to 29 (73%). The main reason for the higher attrition in the control group is that those agents followed the training several weeks later,

Table 4.7: Estimated treatment effect and estimation samples

Treatment dummy	.0773*** (.0214)
Constant	-.9934*** (.0993)
Individual fixed effects	Yes
Control variables	Yes
Observations	1532
Number of agents	55
Number of agents (treatment group)	26
Number of agents (control group)	29
R ²	.6178

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. All standard errors are clustered at the agent level. The samples is defined as all agents minus agents dropping out due to selection. Control variables include *working hours*, *share peak hours*, a linear time trend, and *calls per FTE*.

which increases the probability of dropping out before actually participating in the training. Calculating the attrition rate for treatment group agents who were staying at least until the last week of our full sample which contains longer panel information, we find there is hardly any difference in attrition between the treatment and control groups. In addition, we do not find a systematic pattern in the exit dates that reveals that agents may have left the department due to the upcoming training.¹⁶ Table 4.7 shows that for this sample the estimated treatment effect decreases to 7.7%, which is lower but not significantly different from the estimated effect for the main sample (see Column (3) Table of 4.4).

4.4.3 Effects on call quality

Throughout this chapter, we used a transformation of the average length of calls as a measure of performance. Average handling time provides a clear measure of performance that allows us to quantify the returns to training precisely. One can, however, argue that individual productivity is a function of two dimensions. Calls can differ not only by quantity, but also by quality provided by the agent. In this case, ‘actual’ productivity $y'_{it} = f(y_{it}, y^q_{it})$ is a function of the quantity-performance measure y_{it} and a performance measure for quality y^q_{it} . If y'_{it} is a non-monotonic function in y_{it} , improvements in average handling time will not necessarily translate

¹⁶There are potentially two dates on which selective agents may be more likely to leave due to the training. First, several months ahead of the first training, agents were informed that the whole department was going to be trained. Second, each agent received her schedule four weeks ahead. At neither of the two moments did exit rates appear to be higher than usual.

Table 4.8: The effect of training participation on call quality

Outcome	$\log(y_{it}^q)$ (1)	$\log(y'_{it})$ (2)	Grade 1 (3)	Grade 2 (4)	Grade 3 (5)
Treatment dummy	.0425 (.0391)	.0985*** (.0228)	.8701** (.3721)	1.5742*** (.4439)	-.4701 (.9417)
Constant	-.5157*** (.1374)	-1.2132*** (.0876)	6.0604 (3.8802)	3.2397 (3.6075)	5.2156 (5.4162)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes
Control variables	Yes	Yes	Yes	Yes	Yes
Observations	1730	1730	112	115	107
Number of agents	74	74	14	15	15
R ²	.3384	.5763	.0283	.0474	.0405

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All standard errors are clustered at the agent level. Call quality y_{it}^q is measured as $1/(\text{share of repeat calls} * 10)$. The composite productivity measure y'_{it} is defined as $y'_{it} = aht_{it} * (1 - \text{share of repeat calls})$. Grades are given on 'knowledge of agent' (*Grade 1*), 'agent understood question' (*Grade 2*), and 'solution of the problem' (*Grade 3*). Grades are given on a scale of 1 (very bad) to 10 (very good). Control variables include *working hours*, *share peak hours*, a linear time trend, and *calls per FTE*.

into higher productivity. Since call quality is much more difficult to monitor than work speed, agents may aim to improve their handling times by providing lower call quality. However, agents cannot easily shirk by delivering low quality with a low average handling time because their calls are regularly monitored by their team leaders.

To test whether agents substitute higher performance y_{it} by providing lower quality y_{it}^q , we employ an alternative indicator for which we have data for each worker and week: the share of repeat calls. This measure is defined as the share of customers an agent talked to who called the call centre again within seven days. The firm uses this measure as an indicator of call quality because low customer satisfaction with the first call is the main reason for repeat calls.¹⁷ On average, 23.6% of the customers called back within seven days after the first phone call (see Column (1) of Table 4.1). Because low values of repeat calls (rc_{it}) indicate higher performance, we define call quality as $y_{it}^q = \frac{1}{rc_{it} * 10}$, with an average of 0.454.

The first column of Table 4.8 shows that training participation does not significantly affect call quality. The argument that the shorter average handling time after the training may be associated with lower call quality is therefore not supported by the data.

¹⁷Note that the type (and potentially also the length) of repeat calls can be different from that of initial calls. Repeat calls are, however, not linked to certain agents. This implies that every agent has a given probability of receiving repeat calls.

4 The effects of training on own and co-worker productivity

Having information on the quality of calls allows us to create a composite measure of productivity that incorporates both quantity and quality dimensions. We define y'_{it} as $y'_{it} = y_{it} \cdot (1 - rc_{it})$. Higher shares of repeat calls rc_{it} can thus be interpreted as a penalty in performance, since each additional percentage point in the share of repeat calls (lower quality) relates to a lower composite productivity y'_{it} . The second column of Table 4.8 shows that the estimated treatment effect on the composite measure of productivity is 9.85% and statistically significant at the 1% level.

In addition to the information on the share of repeat calls, we use data gathered in a customer satisfaction survey. Individual calls made by call agents were randomly selected and evaluated by customers on a scale from 1 (very bad) to 10 (very good). Customers gave grades on three different dimensions: the ‘knowledge of the agent’ (*Grade 1*), whether the ‘agent understood the question’ (*Grade 2*), and whether the agent had a ‘solution to the problem’ (*Grade 3*). Because the week t in which the call was made and the corresponding agent i are known, we are able to match this information to whether the agent was treated yet or not. The unit of observation for this analysis is thus a single call, and not agent-week information. Columns (3) through (5) of Table 4.8 indicate show the estimation results. They indicate that the training had a positive effect on the agent’s knowledge and understanding capabilities. However, according to Column (5), the training had no effect on the agent’s ability to provide a solution to the problem. Despite the fact that the analysis using customer-survey data is based on a rather small sample, the results are in line with the previous results showing that quality is not substituted at the cost of average handling time.

4.4.4 Heterogeneous treatment effects

The preceding analysis did not consider interaction effects between worker characteristics and treatment. This section analyses whether an agent’s tenure and number of working hours affect the size of the estimated treatment effect.

First, more experienced agents may exhibit a different effect from the treatment compared to less experienced agents. On average, more experienced workers perform better than less experienced workers. When a ceiling exists in the potential productivity of call centre agents, less experienced agents have higher potential gains from attending the training. In contrast, complementarity in human capital acquisition can lead to greater effects from participation in training for more experienced agents. Column (1) of Table 4.9 shows the treatment effect for agents with a tenure below the median tenure in our sample. Compared to Column (2), which shows regres-

Table 4.9: Direct treatment effect with interaction terms

Sample	Low tenure (1)	High tenure (2)	All (3)
Treatment dummy	.0811*** (.0243)	.0977*** (.0325)	.1640*** (.0378)
Working hours			-.0011 (.0011)
Working hours \times treatment			-.0037*** (.0014)
Constant	-1.0476*** (.0988)	-1.0467*** (.1265)	-1.0027*** (.0872)
Individual fixed effects	Yes	Yes	Yes
Control variables	Yes	Yes	Yes
Observations	908	951	1859
Number of agents	36	38	74
R ²	.1277	.1779	.1585

Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: $\log(y_{it})$. Column (1) and (2) show the baseline regression for agents with below-median (1) and above-median (2) tenure, respectively. All standard errors are clustered at the agent level. Control variables include *share peak hours*, a linear time trend, and *calls per FTE*.

sion results for agents with a tenure above the median, the estimated treatment effect is slightly higher for experienced agents (by 1.7 percentage points). The point estimates, though, are not significantly different from each other.

Second, agents with more working hours may experience a different treatment effect compared to agents with fewer working hours. Column (3) of Table 4.9 shows that the interaction effect between the number of working hours and the treatment dummy is significantly negative. This implies that the training participation of agents with more working hours has a lower effect on their productivity than the training of agents who work fewer hours. This result can be explained by either the greater fatigue of agents with more working hours or the selection of individuals into contracts with lower working hours based on unobservable characteristics such as ability and motivation. Students, for instance, may have contracts with shorter average working hours.

4.5 Conclusions

This chapter analyses the effect of training participation on worker performance by means of a field experiment held at a telephone company call centre. Agents had to participate in a compulsory five-day training programme. We randomly

4 *The effects of training on own and co-worker productivity*

assigned agents to training groups, thereby generating exogenous variation in training participation. Regression results show that the performance of agents was about 10% better after having followed the training. This return is much higher than the returns to training found in studies that attempt to correct for selectivity. This may be explained by the fact that the effects of training on productivity are much greater than the returns measured by workers' wages (Dearden *et al.*, 2006; Konings and Vanormelingen, 2010). However, as mentioned in the introduction, we also expect that returns to training in the call centre sector will be relatively high because the sector lacks proper vocational education for their core jobs. Furthermore, we find that the performance increase is not due to lower quality provided by the worker as indicated by a higher rate of repeat calls. Instead, training participation has a positive effect on customer ratings on agents' knowledge and understanding of the customers' question.

Moreover, we show that agents who did not participate in the training also improved their performance through externalities. These indirect effects may arise due to either knowledge spillover from trained to untrained workers or social pressure. We find that increasing the share of treated peers by 10 percentage points leads to a 0.51% increase in performance. We show that the effects measured are indeed caused by the training and not other effects, such as selective labour turnover.

Our finding of externalities in the workplace has important implications for the estimation of the effects of work-related training on wages using individual data. Even if estimates properly control for selective training participation, the estimated treatment effect is likely to be an underestimation of the true treatment effect, because externalities of training participation increase the productivity of those who have not been trained. Since the tasks carried out by call agents follow an individual production technology without team production, these externalities are probably even higher in other sectors of industry in which team work is more important than in the call centres.

The finding of externalities has important implications for firms' training strategies. When the externalities of training participation increase the productivity of workers who have not been trained, it may be more efficient to design training and the assignment by internalising the treatment externalities. This does, however, involve important questions for further research. Are there nonlinearities in the externalities and what is the optimal share of workers being treated? How and to what extent do externalities propagate further to other workers?

5 Under pressure: Supervisor's role in implicit targets and worker effort

5.1 Introduction

A key question in personnel economics is how to guarantee optimal worker effort.¹ While there is substantial literature on the effects of incentives on performance, knowledge is limited on the role of supervisors in the production process. It is difficult to analyse this role empirically since the mechanisms by which a supervisor can influence worker productivity are not obvious. The overall value-added of supervisors may be the result of different ways of stimulating subordinates, e.g. by monitoring and evaluating, or by motivating workers (Lazear *et al.*, 2012). One strand of this literature analyses the effects of providing relative performance information (rankings) to workers (Blanes i Vidal and Nossol, 2011; Barankay, 2011a,b).

In this chapter, we investigate the role of supervisors in ensuring optimal effort provision on the work floor. In particular, we are interested in a specific mechanism through which supervisor feedback and pressure can push subordinates to provide optimal effort. In our setting, supervisors and workers are regularly informed about individual workers' performance, within team rankings, and the relative performance of their team (team ranking). We argue that supervisors will use these performance rankings to put pressure on underperforming workers.²

We use rich performance information from the personnel records of the call centre of a multi-national telephone company in the Netherlands. These data contain objective performance information for each working week. We use these data to empirically test the influence of supervisors, i.e. the team leaders, on the effort provision of their subordinates. In this particular firm, all workers are organised in teams which are managed by team leaders, and perform the same tasks. Team leaders are responsible to the management of the firm for their teams' overall productivity, and will therefore pressure underperforming workers in their teams by transmitting performance rankings to these workers.

This chapter aims to contribute to the literature on how managers can affect worker performance. Whereas Lazear *et al.* (2012) estimate the overall effect of managers on productivity, which may entail different aspects of a supervisor's work, such as training, monitoring, and motivating subordinates. We focus on one specific channel, namely the effect of performance rankings, by which supervisors encourage

¹This chapter is based on joint work with Ben Kriechel. This chapter has benefitted from useful feedback from Tim Barmby, and participants at the Colloquium on Personnel Economics 2012, the 2012 Annual Conference of the Scottish Economic Society 2012, and the IZA-MIT workshop on leadership 2012.

²In the remainder of this chapter, we use the terms 'motivating' and 'pressuring' interchangeably. We argue that both essentially have the same effect though they may have different side effects, e.g. on job satisfaction.

their subordinates to provide effort to improve their ranking as well as the ranking of the supervisor's team. Supervisors can influence their subordinates' performance by applying pressure. More specifically, we investigate the influence of agents' within-team rankings as well as the influence of the agent's team ranking in the overall team ranking.

We build on the literature on the effects of the revelation of individuals' performance rankings on effort provision. Blanes i Vidal and Nossol (2011) show that the revelation of relative performance information to workers has positive effects on worker productivity. In contrast, Barankay (2011a, 2011b) shows in two field experiments –one among the salespeople of a furniture company and the other among a group of online workers– that the revelation of ranking information can negatively affect future performance for both the extensive and intensive margins. This suggests that at least certain types of workers become less motivated when they have to face information on their relative performance.³

We find that the effect of past individual performance rankings on current performance is much stronger than that of past team rankings on current performance. The effect, however, is strongest among agents who were highest-ranked agents in the previous week while the effect is about half that size for lower-ranked agents. Furthermore, we show that team leader pressure has stronger effects on high ability agents than on low ability agents.

The remainder of the chapter is organized as follows. Section 5.2 introduces the theoretical framework. Section 5.3 describes the data. Section 5.4 describes our empirical approach. Section 5.5 discusses the results and Section 5.6 concludes the chapter.

5.2 Theoretical framework

A simple way of stimulating worker performance is to provide rewards for high performance and punishments for low performance. Firms can use pressure to stimulate workers, especially at the lower end of the performance distribution. This idea is related to a tournament model where workers are punished by being fired (Gürtler

³Studies on the effect of the revelation of rankings on educational outcomes, however, show positive effects. In a field experiment, Azmat and Iriberry (2010) find that high school students who received information about their relative performance performed better than without such information. This result is confirmed by Bandiera *et al.* (2012), who show that a university department that provides information on individual performance (grades) to students increases their performance in subsequent tests. In addition, Tran and Zeckhauser (2012) show that providing information to Vietnamese students on their relative performance significantly increases performance.

and Kräkel, 2012). At the same time, firms can use promotions as an incentive to stimulate effort at the upper end of the performance distribution, even with muted monetary incentives.

In our framework, there are three types of actors: (i) workers who provide the service in the call centre, (ii) supervisors who are responsible for encouraging these workers' effort provision by pressuring workers to provide sufficient effort and, (iii) firm management, who wants both supervisors and workers to provide sufficient effort to make the company profitable.

We assume that individual workers maximise their utility U . Utility is negatively influenced by effort provided, E (i.e., there is a cost of effort), and positively influenced by the wage W an individual receives from work:

$$U = U(E, W) \tag{5.1}$$

with $U'_E < 0$ and $U'_W > 0$.

The worker's output or productivity is influenced by his or her ability A , effort E , and random fluctuations ϕ :

$$y = f(E, A, \phi) \tag{5.2}$$

with $f'_E > 0$ and $f'_A > 0$.

The task is the same across agents i and teams j and can be measured in one dimension.⁴ We assume that workers as well as team leaders are informed about the workers' individual performance, and average performance of their team.

We assume that there are no monetary bonuses; that is, workers are paid fixed wages only. Even with muted explicit incentives firms usually have implicit incentives. Lazear (2000) shows that there is a minimum level of output y_0 that is equal to the minimum output acceptable to the firm.⁵ If workers continuously perform below the minimum level of performance y_0 , the probability of being fired increases. To avoid excessive firing and thus worker turnover, supervisors pressure workers to

⁴This does not preclude other dimensions in which workers' performance can be measured (e.g. quality). However, it is important for *all* workers to be (mainly) evaluated on one measure of performance.

⁵In Lazear's (2000) study, this is the individual worker with A_0 who is indifferent between work and leisure. Because there is competition from outside options, workers compare their wages with those paid by other firms for similar jobs. Lazear (2000) implies that workers of ability between A_0 and A_h work for the firm and effort E depends on ability. In this case, workers of higher ability ($A > A_0$) can provide less effort and still earn rents because they can produce y_0 at lower effort levels than workers with A_0 . If there are no explicit incentives or career concerns, workers will provide the minimum performance level required to keep the job.

perform better. The ranking of workers within teams is a way in which the threat of dismissal is institutionalised and becomes apparent to the worker. This implicit incentive can be formulated as a tournament in which there is no prize for the winner but, rather, the loser is fired with some probability (Lazear and Rosen, 1981).⁶ Similarly, agents can be promoted to supervisor positions, which might explain pressure in the upper tail of the ranking distribution.

Supervisors, however, have incentives to improve overall team performance⁷ and will therefore punish low-performing workers so that all workers provide output $y \geq y_0$. It is difficult for supervisors to measure worker effort provision, that is, to establish suitable values for y_0 (Meyer and Vickers, 1997; Fama, 1980; Holmström, 1999). One way of circumventing this problem is to avoid establishing absolute thresholds and, instead, punish or reward relative worker performance.

Supervisors are evaluated by the management of the organisation.⁸ This evaluation is based on the performance of the workers in their team, potentially conditional on observable worker characteristics. Similarly, supervisors could be evaluated by means of absolute thresholds or team rankings r_j that can be considered a reversed tournament with negative incentives. This holds when the lowest-ranked team leaders have the highest probability of getting fired, conditional on, for example, tenure. The aim of this tournament is to induce workers to work at $y \geq y_0$ and thus improve overall performance of the organisation. In essence, supervisors are involved in a similar tournament as their workers, with the exception that their output depends on their success to elicit output from their subordinate workers:

$$\max_P W_S + [pr(win)|P] \cdot (W_{S1} - W_{S2}) - C(P) \quad (5.3)$$

where $pr(win)$ is the probability of ‘winning’ a reward in the tournament (e.g., a bonus or promotion) and W_{S1} is the winner’s wage level versus the common wage level of supervisors W_{S2} . Supervisors provide effort P for their task of pressuring their subordinate workers, which is supposed to positively influence their productivity. Providing effort P , however, comes at costs $C(P)$.

⁶Gürtler and Kräkel (2012) develop a theoretical model of such dismissal tournaments that uses two levels of ability.

⁷One reason for stipulating that the firm continuously wants to improve overall performance is competition. It suffices that wage increases that are negotiated at a more central level within the firm must be met by productivity increases.

⁸For team leader incentives, see Bandiera *et al.* (2007).

5 Under pressure: Supervisor's role in implicit targets and worker effort

This implies a first-order condition for the supervisor:

$$\frac{\partial[\text{pr}(\text{win})]}{\partial P} \cdot (W_{S1} - W_{S2}) = C' \quad (5.4)$$

The output of the supervisor is dependent on his or her subordinates:

$$y_S = \sum_i f_i(E_i, \phi_i | P) = \sum_i (E_i | P) + \theta_i \quad (5.5)$$

where θ_i reflects the random fluctuation in output of an individual supervisor i .⁹

Similar to the workers, supervisors compete for the best position. In a two-supervisor setting, with each supervisor $i = [1, 2]$ identified by suffixes, Supervisor 1 wins the tournament as long as $y_{S1} > y_{S2}$; that is,

$$\text{Pr}(y_{S1}) > \text{Pr}(y_{S2}) \quad (5.6)$$

$$\Leftrightarrow \text{Pr}(P_1 - P_2 > \theta_1 - \theta_2) > 0 \quad (5.7)$$

$$\Leftrightarrow H(P_1 - P_2) > 0 \quad (5.8)$$

with $H(\cdot)$ denoting the cumulative distribution of the disturbance θ , which is assumed to have zero mean and variance $2\sigma^2$. The supervisor's response is then

$$(W_{S1} - W_{S2}) \cdot h(P_1 - P_2) = C'(P_1) \quad (5.9)$$

where $h(\cdot)$ is the density function of the disturbance. Assuming symmetry, that is, $P_1 = P_2$, this becomes

$$(W_{S1} - W_{S2}) \cdot h(0) = C'(P_1) \quad (5.10)$$

The supervisor is essentially competing in a similar type of tournament as the workers, the only difference being the prizes, error structure, and production function. The underlying logic is, however, the same.

The model implies that relative performance measurement is important. Given that all team leaders manage their own team, they are most likely interested in their subordinates' ranking within the team, whereas the team's overall ranking, in turn, translates into pressure on the team leader.

⁹The random fluctuation is the sum of all identical and independently distributed fluctuations of the subordinate workers.

Table 5.1: Descriptive statistics

	mean	s.d.	min	max
Performance (y_{it})	0.3111	0.0820	0.1606	0.6219
Residual performance ($\hat{\varepsilon}_{it}$)	0.0004	0.0036	-0.0103	0.0678
Tenure (in years)	1.528	2.872	0.019	12.273
Gender (1=male)	0.3278	0.4700	0	1
Age (in years)	29.3	10.1	17.5	65.3
Agent rank in team $t - 1$	0.4239	0.2398	0	0.9767
Team rank $t - 1$	0.4063	0.2260	0	1
Number of observed weeks	63	32	1	116

Descriptive statistics are calculated at the mean of each agent. The overall number of agents is 418 with a total of 14,905 agent-week observations. Performance y_{it} is defined as the inverse of average handling time (ah_{it}), multiplied by 100: $y_{it} = \frac{1}{ah_{it}} \cdot 100$. Residual performance $\hat{\varepsilon}_{it}$ is defined as the residual of a regression of performance y_{it} on log-tenure, week fixed effects, and individual fixed effects, cleaned from random fluctuation (cf. Section 5.4). *Agent rank in team*, *agent rank in department*, and *team rank* are defined as 1 if the agent (team) is the best-performing agent (team) in week t , and 0 if worst.

5.3 The firm

To analyse the effects of pressure on worker performance, we use weekly information on the performance of individual workers. We use data from the largest department, that for private customers with fixed cell phone contracts. In total, we observe 418 agents who are monitored and evaluated by 16 team leaders. The sample spans a period of 2.44 years. The agents are, on average, 29 years old and have an average tenure of 1.5 years (see Table 5.1).

The firm uses different incentive mechanisms to reward agents who outperform their peers. First, all agents are evaluated by their team leaders in an annual appraisal interview that determines both the annual wage increase as well as an annual bonus. For both, the firm's management sets reference values that are then adjusted according to the individual appraisal. The appraisal grade, measured on a scale from one (lowest) to five (highest), comprises several measurable performance outcomes, such as average handling time, measures of customer satisfaction, as well as the agent's behaviour towards peers, team leaders, and managers in the previous year. Higher salary increases and annual bonuses are only ensured for workers who outperform their peers. The annual wage increase can be up to 8%; the average annual wage bonuses is about 6-8%, with a maximum of about 12%.

Second, supervisors are regularly recruited from within the worker pool and it speaks for itself that worker productivity, as a sign of ability and willingness to provide effort, is important measures in determining these promotions. While pro-

motions are a potential incentive mechanism, there is only a low number of team leaders relative to the overall number of agents.¹⁰

The agents' task is to handle the inbound calls of the company's customers, who contact the call centre for problems, complaints, or information. This task consists of talking to customers and making notes in the firm's customer database. Performance is measured by the main key performance indicator used by the firm to evaluate call agents: average handling time. This measure is defined as the average time an agent needs to handle a customer call during one calendar week (cf. Liu and Batt, 2007; Chapters 3 and 4).

The department's aim is to improve performance by decreasing average handling time. This chapter uses the inverse of the average handling time (aht_{it}) multiplied by 100: $y_{it} = \frac{1}{aht_{it}} \cdot 100$. Since lower values of aht_{it} are interpreted as higher productivity, this transformation allows us to interpret increases in y_{it} as increases in performance. The average productivity of all agents in our sample is 0.311, which translates to 5 minutes and 35 seconds for an average call.

There is substantial heterogeneity in individual productivity within and between agents. This suggests that not only individual-specific characteristics, but also other, department-specific effects such as technical problems matter for the individual productivity of all agents working in the department. In a regression of performance on worker-specific fixed effects only, 48% of the overall variation can be explained, while week fixed effects explain only 7%.

In general, incoming customer calls are assigned randomly to agents. Agents have no direct influence on the types of calls they receive or the types of customers put through to them and therefore have the same probability of exceptionally long or short calls. Similarly, there is no specialisation of teams in certain types of calls. Because neither managers nor team leaders have any influence on the types of calls, agents as well as teams can be ranked by average handling time.

5.4 Empirical approach

The implication of the tournament model presented in Section 5.2 is that agents are likely to receive more pressure if they are either low-ranked within their team or if they work in a team with low average performance. To test these implications

¹⁰Chapter 6 reports evidence on the effects of the introduction of explicit incentives on performance outcomes using data from the same call centre. The introduction of explicit incentives based on observable performance outcomes does not overlap with the observation period in this chapter.

empirically, we proxy supervisory pressure by the ranking of the worker's team. The ranking variables are used to explain the outcome variable, worker-week performance.

The reduced form of this model is

$$y_{it} = \alpha_i + \gamma_t + \log(d_{it})'\beta_1 + r'_{i,t-1}\theta_1 + r'_{j,t-1}\theta_2 + v_{it} \quad (5.11)$$

where the outcome variable y_{it} is the individual performance of agent i in week t ; α_i and γ_t are the fixed effects for each worker and each week, respectively; and d_{it} is individual tenure. The variables $r_{i,t-1}$ and $r_{j,t-1}$ are the rank of a worker in his or her team and the ranking of the worker's team compared to other teams, respectively. Both variables $r_{i,t-1}$ and $r_{j,t-1}$ are defined on a scale from zero to one, where the lowest-ranked agent (team) is scored a zero and the highest-ranked agent (team) a one.

Instead of estimating Equation (5.11) directly, we apply a two-step approach. In the first step, we regress individual performance y_{it} on individual and time fixed effects and tenure:

$$y_{it} = \alpha_i + \gamma_t + \log(d_{it})'\beta_1 + \varepsilon_{it} \quad (5.12)$$

where all variables are defined as above. The idiosyncratic error term ε_{it} in Equation (5.12) entails purely random fluctuations, as well as variation due to supervisory pressure.

The second step uses the error term $\hat{\varepsilon}_{it} = \hat{y}_{it} - y_{it}$ to estimate the effects of individual and team rankings on the variation that is unexplained by the other covariates. This two-step approach allows us to net out variation in individual performance from time-constant individual effects, week-specific effects, and time effects. This approach also has the advantage of reducing the risk of serial correlation, since time-invariant effects are cancelled out by the individual fixed effects in Equation (5.12).

To estimate the effect of rankings on individual performance, we regress $\hat{\varepsilon}_{it}$ on information on the previous period's rank of a worker in the team, $r_{i,t-1}$, and the rank of the worker's team, $r_{j,t-1}$:

$$\hat{\varepsilon}_{it} = a + r'_{i,t-1}b_1 + r'_{j,t-1}b_2 + \nu_{it} \quad (5.13)$$

This equation implicitly assumes a symmetric effect for low- and high-performing agents and teams. To relax this assumption, b_1 and b_2 are estimated separately for below- and above-median agents and teams.

The main challenge of this empirical approach is that the error term $\hat{\varepsilon}_{it}$ contains both random variation (or regression to the mean) and variation due to pressure (which is unobserved to the econometrician). When Equation (5.13) is estimated without correcting the error term $\hat{\varepsilon}_{it}$ for random variation, the effect of the rankings is likely to be overestimated due to regression to the mean. We therefore propose to net out the error term by the natural fluctuation in performance (Barnett *et al.*, 2005). This procedure calculates a regression to the mean effect as

$$r = \frac{\sigma_w^2}{\sqrt{\sigma_w^2 + \sigma_b^2}} C(z) \quad (5.14)$$

$$= \sigma_t(1 - \rho)C(z) \quad (5.15)$$

where $\sigma_t^2 = \sigma_w^2 + \sigma_b^2$ is the total variance, $\sigma_w^2 = (1 - \rho)\sigma_t^2$ is the within-subject (agent) variance, $\sigma_b^2 = \rho\sigma_t^2$ is the between-subject variance, and ρ is the correlation. The term $C(z)$ is defined as $C(z) = \phi(z)/\Phi(z)$, where $\phi(\cdot)$ is the probability density and $\Phi(\cdot)$ is the cumulative distribution function. The variable z is defined as $z = (c - \mu)\sigma_t$ for observations above c and as $z = (\mu - c)\sigma_t$ for observations below c , where μ is the population mean.

Since the data contain multiple observations of the same persons, we can calculate r_{it} and net out $\hat{\varepsilon}_{it}$ from random variation that may be due to regression to the mean. Since the remaining fluctuations in the second step are moderate, these adjustments to correct for the regression to the mean are moderate.¹¹

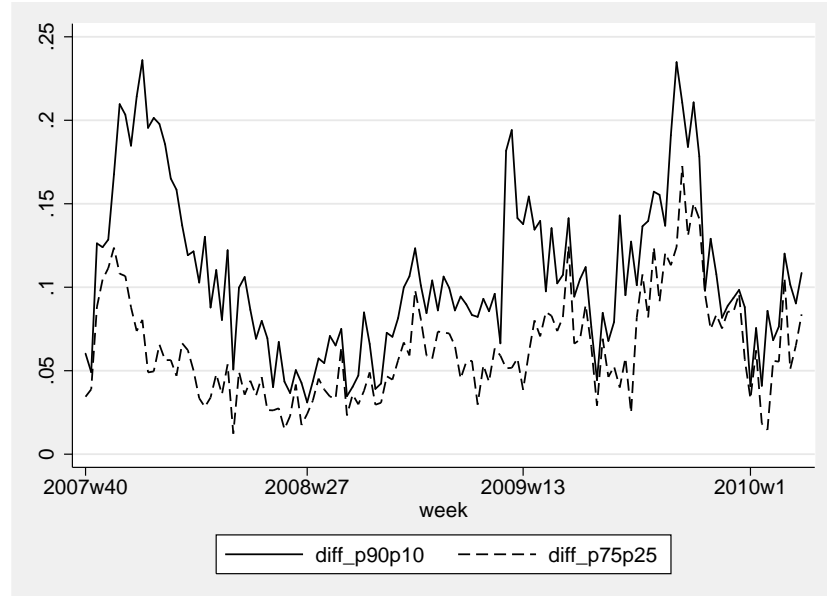
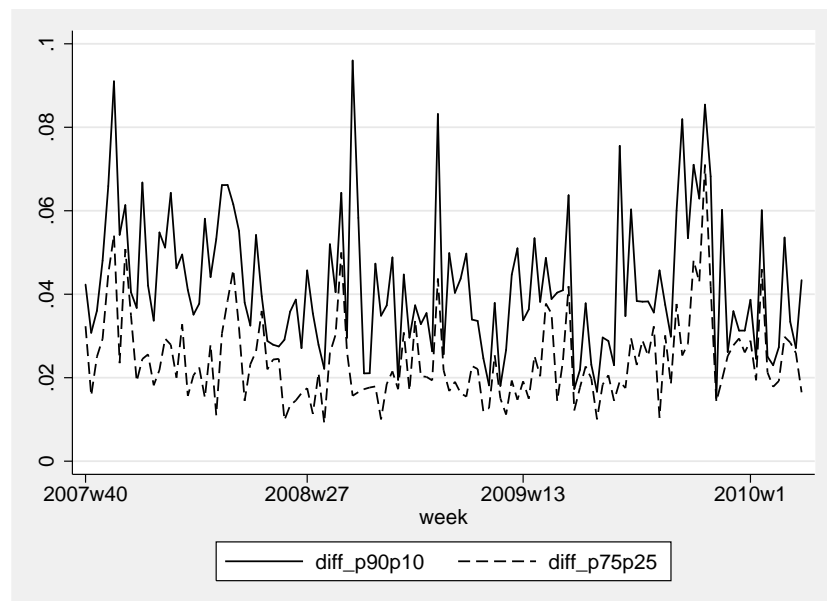
5.5 Results

5.5.1 Descriptive evidence

Panels (a) and (b) of Figure 5.1 shows the difference in absolute performance between teams at the upper and lower ends (percentiles) of the performance distribution for performance y_{it} and residual performance, that is, the performance net of regression to the mean (see the previous section), $\hat{\varepsilon}_{it}$. The solid (dashed) line shows the absolute difference in average team performance in week t , $\bar{y}_{jt} = \frac{1}{N_j} \sum_{i=1}^{N_j} y_{it}$, of the teams at the 90th (75th) and 10th (25th) percentiles.¹²

¹¹We offer an alternative, a one-step estimation in which the rank positions of the workers and their team are estimated along with all control variables from the first step's approach. As discussed in the next section, the outcome is in line with the two-step approach with correction for the regression to the mean.

¹²Note that the percentiles are defined for each week separately; the teams compared in Panels (a) and (b) of Figure 5.1 may therefore change from week to week.

Figure 5.1: Percentile differences in performance y_{jt} and residual performance $\hat{\varepsilon}_{it}$ (a) Performance y_{jt} (b) Residual performance $\hat{\varepsilon}_{it}$

Note: diff_p90p10 (diff_p75p25) is the absolute difference between the average performance y_{jt} (Panel (a)) and residual performance $\hat{\varepsilon}_{it}$ (Panel (b)) in the 90th (75th) percentile and the team in the 10th (25th) percentile of the ranking in week t . The weekly ranking is defined on a scale from 0 (lowest ranked team of team j in week t) to 1 (highest ranked team).

5 Under pressure: Supervisor's role in implicit targets and worker effort

The figure shows that there are substantial differences in performance between teams at the upper end of the performance distribution and those at the lower end. This suggests variation between well-performing teams and teams that perform poorly. Furthermore, the figure shows substantial variation in relative performance over time. Though this may be the result of other determinants, such as random fluctuations, this may be caused by supervisory pressure, captured by the feedback supervisors give to their subordinates on these team rankings.

The difference between low- and high-performing teams can also be examined by testing whether it is significantly different from zero. We therefore test whether the average performance of the high-performing teams—at the 90th and 75th percentiles—is statistically different from the average performance of the low-performing teams—at the 10th and 25th percentiles. For all combinations of teams, the better teams perform significantly better than the worst performing team of a week.

5.5.2 Estimation results

The effects of the workers' and their teams' ranking are shown in several types of estimations. First, we establish the influence of lags of worker rank positions as well as team rank positions on the unexplained variation in performance, as well as a combination of both. Second, we show that the effect of a worker's rank position is not linear but, instead, increases with lagged performance. The outcome variable for these estimations is residual performance that is furthermore netted from regression-to-the-mean effects (see Section 5.4 of this chapter).

The effect of agent and team ranking on individual performance

Table 5.2 shows the results estimating the influence of a call agent's rank position in a team for each of the previous weeks on the current week's effort. The simple ordinary least squares (OLS) regression shows a positive relation between the overall rank of a worker in the previous week (*agent rank in team $t - 1$*) with the current unexplained variation in productivity. Adding lags to the previous weeks $t - 2$, $t - 3$, $t - 4$, and $t - 5$ for the two-, three-, four-, and five-week lags, respectively, shows that only the first and second lags have a significant positive effect on current productivity. This finding suggests that the previous week's rank has an especially important effect on an agent's current performance.

Table 5.3 repeats this estimation for the rank position of a call agent's team relative to other teams in the department. Here again, only the first lag of a week has a positive influence on an individual worker's productivity. The team effect on an

Table 5.2: OLS results of previous individual ranking

	(1)	(2)	(3)	(4)	(5)
Agent rank in team ($t - 1$)	0.036 (0.002)***	0.031 (0.003)***	0.031 (0.003)***	0.032 (0.003)***	0.033 (0.003)***
Agent rank in team ($t - 2$)		0.008 (0.002)***	0.008 (0.003)***	0.009 (0.003)***	0.011 (0.003)***
Agent rank in team ($t - 3$)			-0.000 (0.003)	0.001 (0.003)	0.004 (0.003)
Agent rank in team ($t - 4$)				-0.004 (0.003)	-0.000 (0.003)
Agent rank in team ($t - 5$)					-0.012 (0.003)***
Constant	-0.018 (0.001)***	-0.020 (0.002)***	-0.020 (0.002)***	-0.019 (0.002)***	-0.018 (0.002)***
R^2	0.03	0.04	0.04	0.03	0.03
N (agents)	410	404	396	384	375
N	14,487	14,077	13,673	13,277	12,893

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: Residual performance $\hat{\varepsilon}_{it}$ (see Equation (5.13)). $\hat{\varepsilon}_{it}$ is defined as the residual of a regression of performance y_{it} on log-tenure, week fixed effects, and individual fixed effects, cleaned from random fluctuation (cf. Section 5.4). *Agent rank in team* is defined as 1 if the agent is the best-performing agent in team j and week t , and 0 if worst. All regressions are clustered at the agent level.

agent's current performance, however, is only about half as strong as the effect of the individual agent's ranking.

The third set of OLS regressions combines both the rank measure of a worker within a team and the rank of the team (see Table 5.4). Both previous results still hold; the effects of a good team and previous good individual rankings seem to have a positive influence on productivity. This does not imply any effect of pressure on badly performing workers, since only good teams influence the performance of their team members positively.¹³

¹³This result hints at an effect of spillover from teams, as noted by Mas and Moretti (2009), and a tournament for higher positions, as mentioned by Blanes i Vidal and Nossol (2011).

Table 5.3: OLS results of previous team ranking

	(1)	(2)	(3)	(4)	(5)
Team rank	0.015	0.014	0.013	0.014	0.014
($t - 1$)	(0.002)***	(0.003)***	(0.003)***	(0.003)***	(0.003)***
Team rank		0.002	0.001	0.002	0.003
($t - 2$)		(0.002)	(0.002)	(0.002)	(0.003)
Team rank			0.002	0.002	0.004
($t - 3$)			(0.003)	(0.003)	(0.003)
Team rank				-0.004	-0.002
($t - 4$)				(0.002)	(0.002)
Team rank					-0.006
($t - 5$)					(0.002)**
Constant	-0.007	-0.007	-0.008	-0.007	-0.006
	(0.001)***	(0.001)***	(0.001)***	(0.002)***	(0.002)***
R^2	0.01	0.01	0.01	0.01	0.00
N (agents)	410	404	396	384	375
N	14,487	14,077	13,673	13,277	12,893

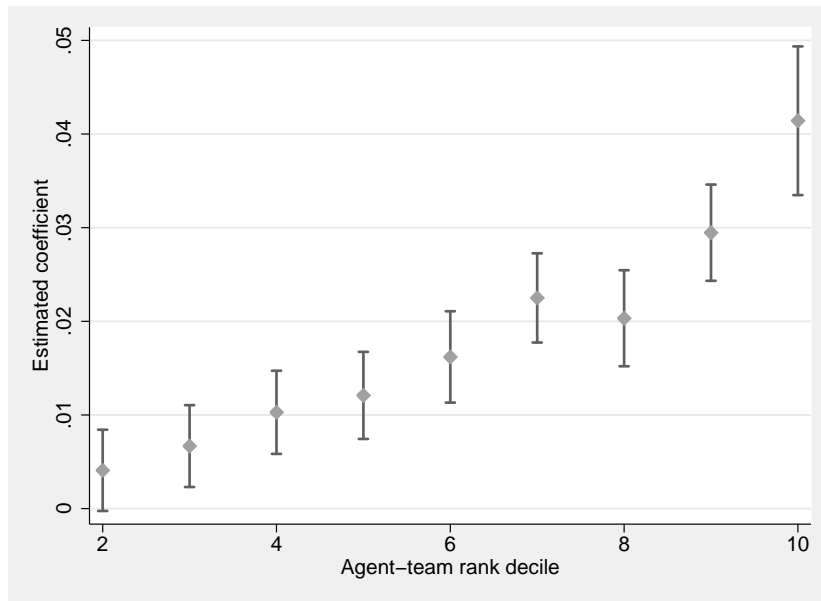
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: Residual performance $\hat{\varepsilon}_{it}$ (see Equation (5.13)). $\hat{\varepsilon}_{it}$ is defined as the residual of a regression of performance y_{it} on log-tenure, week fixed effects, and individual fixed effects, cleaned from random fluctuation (cf. Section 5.4). Team rank is defined as 1 if the agent's team is the best-performing team in week t , and 0 if worst. All regressions are clustered at the agent level.

Table 5.4: Joint estimation of individual and team ranking

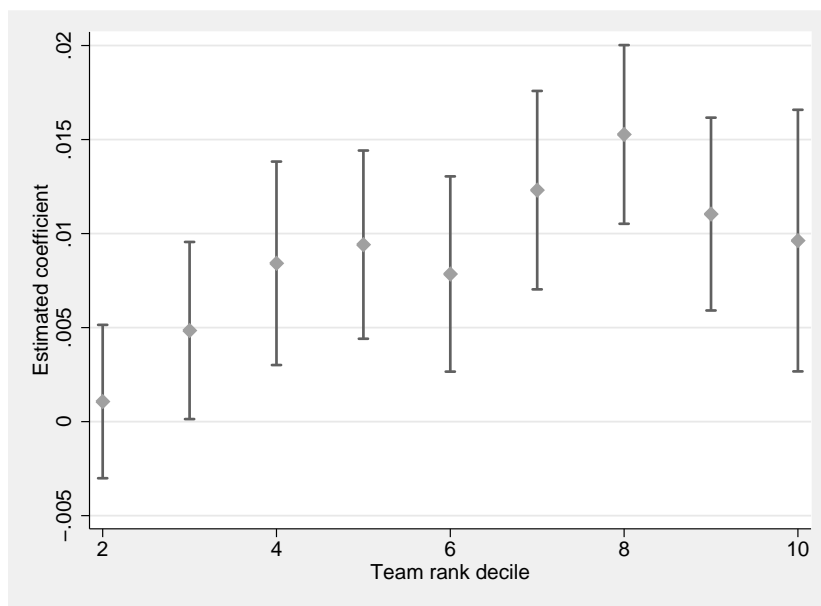
	(1)	(2)	(3)
Agent rank in team	0.036	0.032	0.032
($t - 1$)	(0.002)***	(0.003)***	(0.003)***
Team rank	0.015	0.014	0.014
($t - 1$)	(0.002)***	(0.003)***	(0.003)***
Agent rank in team		0.007	0.008
($t - 2$)		(0.002)***	(0.003)***
Team rank		0.001	0.001
($t - 2$)		(0.002)	(0.002)
Agent rank in team			-0.001
($t - 3$)			(0.003)
Team rank			0.002
($t - 3$)			(0.003)
Constant	-0.025	-0.027	-0.028
	(0.002)***	(0.002)***	(0.002)***
R^2	0.04	0.04	0.04
N (agent)	410	404	396
N	14,487	14,077	13,673

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: Residual performance $\hat{\varepsilon}_{it}$ (see Equation (5.13)). $\hat{\varepsilon}_{it}$ is defined as the residual of a regression of performance y_{it} on log-tenure, week fixed effects, and individual fixed effects, cleaned from random fluctuation (cf. Section 5.4). *Agent rank in team*, and *team rank* are defined as 1 if the agent (team) is the best-performing agent (team) in week t , and 0 if worst. All regressions are clustered at the agent level.

Figure 5.2: Estimated effect of agent-team and team rankings by decile



(a) Agent-team rankings



(b) Team rankings

Note: The figure shows the point estimates and corresponding 95% confidence intervals of dummy variables that are 1 if an agent was in the x th decile in the previous week, and 0 otherwise. Estimates and standard errors are taken from a regression of residual performance $\hat{\varepsilon}_{it}$ on dummies for agent-team-rank (Panel (a)) and team-rank (Panel (b)) deciles 2 (second lowest rank decile) to 10 (highest rank decile).

Heterogeneous effects for low- and high-ranked agents

While the serial correlation of more productive workers and positive spillover effects from other team members may positively influence good workers, there may still be an effect on the lower end of the rank distribution. Panel (a) of Figure 5.2 shows the point estimates of dummy variables for agents who were in the x th decile in the previous week. This approach to assessing the effect of the previous week's ranking on current performance confirms the positive relation shown in Table 5.2 and shows that the effect is considerably stronger for the highest-ranked agents. The illustration of the effect of ranking of teams, however, is much smaller and less significant. This is also confirmed by Panel (b) of Figure 5.2, which shows that the ranking effect increases less with the rank decile.

To investigate this issue more carefully, we apply a spline estimation, using two and five performance segments, respectively, in our estimation model. *Agent spline 1* (*Agent spline 2*) is defined as a linear rank variable for agents below (above) the median rank. The spline variables using five segments are calculated accordingly.

Table 5.5: Spline estimation of the relation between ranking and effort – 1

	(1)	(2)	(3)
Agent spline 1	0.026 (0.005)***		0.026 (0.005)***
Agent spline 2	0.046 (0.007)***		0.046 (0.007)***
Team spline 1		0.025 (0.005)***	0.025 (0.005)***
Team spline 2		0.004 (0.005)	0.004 (0.005)
Constant	-0.016 (0.001)***	-0.010 (0.002)***	-0.026 (0.002)***
R^2	0.04	0.01	0.04
N (agents)	410	410	410
N	14,487	14,487	14,487

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: Residual performance $\hat{\varepsilon}_{it}$ (see Equation (5.13)). $\hat{\varepsilon}_{it}$ is defined as the residual of a regression of performance y_{it} on log-tenure, week fixed effects, and individual fixed effects, cleaned from random fluctuation (cf. Section 5.4). Splines for both agents and teams are equally sized. All regressions are clustered at the agent level.

Column (1) of Table 5.5 shows a regression of residual performance on two splines, where *Agent spline 1* can be interpreted as the original rank variable for agents below the median and *Agent spline 2* as the original rank variable for agents above

5 Under pressure: Supervisor's role in implicit targets and worker effort

the median. In line with the results of Figure 5.2, the estimation results in Table 5.5 show that the effect on the upper half of the ranking (*Agent spline 2*) is almost double the effect for lower-ranked agents. The two estimates are significantly different from each other.

Column (2) of Table 5.5 shows a different picture for the effect of team rankings on an agent's performance: Here, the effect of the team ranking is only significant for teams in the lower half of the performance distribution and not for teams that perform, on average, above the median. Column (3) of Table 5.5 confirms the results in a joint estimation of agent and team splines.

Table 5.6 replicates the results from Table 5.5 with five splines, thus allowing for more flexibility. The results largely confirm the earlier findings. Column (1) shows that the effect is strongest for the highest quintile. At the same time, the effect is insignificant for *Agent spline 4*. This suggests that team leader pressure works somewhat at the far upper end, for example, through positive incentives. The results for the team ranks (Columns (2) and (3)) show that the effects are strongest for those teams somewhat in the middle of the rankings (*Team spline 2*, *Team spline 4*).

In a final regression, we divide the call agents into quintiles to investigate whether pressure is exerted differently on low- and high-performance workers. These quintiles are based on previous performance and reflect the overall ability of a worker. In the past, the lowest (highest) quintile has shown the lowest (highest) productivity. By restricting the estimation to these groups of low (high) performers, we investigate how the performance of agents is influenced given their past performance or ability. The quintiles are calculated as the average ranking from $t - 2$ to $t - 5$. To investigate the heterogeneous effects of individual rankings and team rankings, we separate the rankings in above and below median splines.

Table 5.7 shows two distinctive patterns. First, the results show that higher ability agents react stronger to individual rankings (agent spline 1 / Column (4); agent spline 2 / Column (5)). This result is in line with the earlier results that showed that team leaders rather put pressure on the upper end of the ranking distribution. Second, the results show that the pressure from team rankings is only significant for teams that are low performing. This suggests that managers pressure these team leaders stronger than well performing team leaders. Again, team leaders put through this pressure strongest to high ability agents.

Table 5.6: Spline estimation of the relation between ranking and effort – 2

	(1)	(2)	(3)
Agent spline 1	0.032 (0.013)**		0.032 (0.012)***
Agent spline 2	0.027 (0.014)**		0.027 (0.014)*
Agent spline 3	0.036 (0.014)***		0.036 (0.014)***
Agent spline 4	0.022 (0.014)		0.022 (0.014)
Agent spline 5	0.086 (0.021)***		0.085 (0.021)***
Team spline 1		0.017 (0.014)	0.017 (0.014)
Team spline 2		0.030 (0.014)**	0.030 (0.014)**
Team spline 3		0.012 (0.014)	0.012 (0.013)
Team spline 4		0.022 (0.012)*	0.022 (0.012)*
Team spline 5		-0.024 (0.016)	-0.024 (0.015)
Constant	-0.017 (0.002)***	-0.009 (0.002)***	-0.026 (0.003)***
R^2	0.04	0.01	0.04
N (agents)	410	410	410
N	14,487	14,487	14,487

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: Residual performance $\hat{\varepsilon}_{it}$ (see Equation (5.13)). $\hat{\varepsilon}_{it}$ is defined as the residual of a regression of performance y_{it} on log-tenure, week fixed effects, and individual fixed effects, cleaned from random fluctuation (cf. Section 5.4). Splines for both agents and teams are equally sized. All regressions are clustered at the agent level.

Table 5.7: Spline estimation by past performance quintiles

Quintiles	(1) worst	(2)	(3)	(4)	(5) best
Agent spline 1	0.032 (0.007)***	0.036 (0.008)***	0.031 (0.011)***	0.067 (0.016)***	0.032 (0.026)
Agent spline 2	0.034 (0.029)	0.019 (0.013)	0.020 (0.010)**	0.032 (0.009)***	0.078 (0.016)***
Team spline 1	0.005 (0.008)	0.024 (0.009)***	0.019 (0.008)**	0.026 (0.010)***	0.059 (0.014)***
Team spline 2	0.006 (0.008)	0.014 (0.010)	0.005 (0.008)	-0.005 (0.010)	0.001 (0.014)
Constant	-0.019 (0.003)***	-0.028 (0.004)***	-0.024 (0.005)***	-0.042 (0.008)***	-0.050 (0.012)***
R^2	0.02	0.03	0.01	0.03	0.05
N (agents)	410	410	410	410	410
N	2,814	2,822	2,818	2,812	2,811

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: Residual performance $\hat{\varepsilon}_{it}$ (see Equation (5.13)). $\hat{\varepsilon}_{it}$ is defined as the residual of a regression of performance y_{it} on log-tenure, week fixed effects, and individual fixed effects, cleaned from random fluctuation (cf. Section 5.4). Splines for both agents and teams are equally sized. All regressions are clustered at the agent level.

One-step estimation

Throughout all regressions, we have applied a two-step procedure to net out the performance measures from regression to the mean. This procedure can be seen as a conservative estimate of the effect of pressure on workers. We try to net out as much variation as possible by including week and person fixed effects and tenure in a first step, without allowing the rank of a worker to pick up any of these variations. A straightforward alternative would be to estimate both steps in one estimation that includes all the variables (and fixed effects) of the first step, as well as the variables that indicate the influence of past rank, that is, expected supervisory pressure.¹⁴ The results, given in Table 5.8, are broadly in line with our previous findings. The effects, however, are stronger than shown in previous results. This suggests that the netting out procedure proposed in this chapter indeed reduces bias due to regression to the mean.

¹⁴The results from the first stage are shown in Table 5.9.

Table 5.8: One-step estimation

	(1)	(2)	(3)	(4)
Rank of agent ($t - 1$)	0.075 (0.005)***			
Rank of team ($t - 1$)	0.036 (0.004)***			
Agent spline 1		0.059 (0.006)***		0.068 (0.006)***
Agent spline 2		0.078 (0.010)***		0.084 (0.010)***
Team spline 1			0.040 (0.008)***	0.059 (0.007)***
Team spline 2			0.008 (0.007)	0.015 (0.006)**
log(tenure)	0.035 (0.003)***	0.041 (0.003)***	0.044 (0.003)***	0.034 (0.003)***
Constant	0.150 (0.016)***	0.151 (0.017)***	0.152 (0.018)***	0.149 (0.016)***
Individual fixed-effects	yes	yes	yes	yes
Week fixed-effects	yes	yes	yes	yes
R^2	0.64	0.63	0.61	0.64
N (agents)	410	410	410	410
N	14,487	14,487	14,487	14,487

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: Residual performance $\hat{\varepsilon}_{it}$ (see Equation (5.13)). $\hat{\varepsilon}_{it}$ is defined as the residual of a regression of performance y_{it} on log-tenure, week fixed effects, and individual fixed effects, cleaned from random fluctuation (cf. Section 5.4). Splines for both agents and teams are equally sized. All regressions are clustered at the agent level.

5.6 Conclusion

This chapter analyses the extent to which team leaders can influence worker productivity by exerting pressure on their subordinates who are performing relatively poorly. Team leaders can influence effort provision in environments where effort cannot be easily observed. We propose that the division of independently working agents into independent teams allows for the provision and evaluation of productivity. This setup can be used to compare performance across teams to introduce a means of eliciting supervisory pressure and thus optimal effort provision by agents. By ranking the call agents within teams, team leaders can influence their effort provision. Team leaders are likely to increase pressure on low-performing call agents to increase their performance and thus maintain high average team performance.

Table 5.9: First-stage regression

	(1)
log(tenure, in weeks)	0.044 (0.002)***
Constant	0.198 (0.009)***
R^2	0.62
N (agents)	437
N	16,110

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: y_{it} . The regression controls for individual and week fixed effects. All regressions are clustered at the agent level.

We show that team leader pressure influences high-ranked workers, who are likely to increase their performance in subsequent weeks. The pressure is especially felt among high-performing workers. We can demonstrate this by dividing the call agents into types who were lower- and higher-performing in the past. The estimation results of this analysis show that high ability agents react much more strongly to their individual ranking than low ability agents.

We furthermore show that the effects of individual rankings is much more important than the effects of team rankings. Team rankings, however, are only relevant for lower ranked teams. For higher ranked teams, the team ranking does not have a significant effect on individual performance.

These results imply that even in environments where workers perform independent tasks, the division of workers into teams has the advantage that supervisors encourage (at least certain types of) workers to increase their effort provided.

6 The heterogeneous effects of bonus pay on performance outcomes

6.1 Introduction

Firms use performance pay to link workers' wages to their performance to elicit optimal effort levels. There has been substantial research on measuring the effect of provision of performance bonuses on performance outcomes that depend on either absolute performance targets or workers' performance relative to their peers. Most studies use personnel data and exploit (quasi-) exogenous variations in the compensation schemes of single firms. Overall, studies find that pay for performance based on absolute performance thresholds yields higher worker performance (see, e.g., Dohmen and Falk, 2011, for a recent overview).¹

In this chapter, I exploit the introduction of performance bonuses in the call centre of a multi-national telephone company located in the Netherlands. The dataset contains information on different performance outcomes, one of which was used to incentivise call agents by introducing a bonus payment upon achieving a pre-defined performance level over a given time period. This outcome, which is available for each working week, is based on the service quality an agent provides to customers. This net promoter score is rated by the firm's customers (see, e.g., Keiningham *et al.*, 2007). In accordance with the related literature, I find that the workers, on average, react to the newly introduced performance pay by providing higher performance. This result is also in line with findings from other laboratory as well as field studies for other industries (Lazear, 2000; Shearer, 2004; Shi, 2010; Heywood *et al.*, 2011). Estimation results show that workers who are lower in the performance distribution before the introduction of the performance pay drive the performance improvement after the bonus is introduced, whereas the bonus does not affect the behaviour of workers who would have outperformed the performance target anyhow. These results are stable to tests for seasonality and time trends and other robustness tests.

This chapter contributes to the literature on the analysis of performance incentives on performance outcomes in several ways. First, it contributes to the literature on the effects of introducing a performance-related bonus on performance outcomes by focusing on performance pay related to the quality of workers' output instead of the quantity of output. Previous research exploiting firm personnel data shows that performance-related pay can have substantial effects on individual performance. Lazear (2000) shows that the output of windshield installers increases by 44% after the introduction of piece-rate pay. Similarly, Shearer (2004) and Shi (2010) find

¹A number of related studies examine the effect of relative incentives on performance (Bandiera *et al.*, 2005) and the effect of team incentives on productivity (Hamilton *et al.*, 2003; Muralidharan and Sundararaman, 2011; Bandiera *et al.*, 2012).

that workers perform about 20% better under piece-rate pay in tree planting and tree thinning, respectively. Heywood *et al.*, (2011) find that a piece rate leads to an increase of 50% in peer-reviewed publishing among professors of a university in China. These studies focus on output measures, such as the number of pieces accomplished within a given time. It is usually much more difficult, however, for firms to monitor the quality of workers' output. The call centre analysed in this chapter applies a monitoring system for service quality. Given the availability of this performance measure, the firm incentivised this measure to improve overall call quality.

Second, we analyse whether agents throughout the ability distribution react differently to the incentives set by management. For relative incentives, Azmat and Iriberry (2010) find that ability matters for the response to incentives. In this chapter, I find that the incentives work best for those in the lower tail of the ability distribution, while they decrease performance at the upper end of the distribution.

Third, this chapter analyses the effect on incentivised as well as non-incentivised performance measures. If workers' task performance consist of several dimensions, one of which is incentivised while the others are not, workers may simply aim at performing well on the performance outcome that is rewarded (Holstrom and Milgrom, 1991). Only a few studies analyse the effect on alternative performance outcomes that are not subject to the incentive introduced (Asch, 1990; Al-Ubaydli *et al.*, 2008). This chapter is able to analyse whether an increase in the incentivised service quality comes at the cost of a lower service quantity. However, I do not find such a trade-off, since service quantity is also affected positively.

The following section describes the firm whose performance pay is evaluated and how worker performance is measured. Furthermore, it discusses how the firm set wages before and since the introduction of performance pay. The main estimation results are given in Section 6.3. Section 6.4 shows that the results are robust to alternative hypotheses (e.g., placebo treatments). Section 6.5 summarises and concludes the chapter.

6.2 The firm and its incentive schemes

6.2.1 The firm

The call centre in which this incentive scheme was implemented is part of a multinational telephone company located in the Netherlands.² The call centre is an in-house service centre that handles the inbound calls of current and prospective customers. Customers can contact the call centre to ask questions and report technical and administrative problems and complaints. An automated routing system connects customers to available agents.

The call centre is organised into different departments, some of which were subject to the introduction of performance bonuses based on individual performance.³ The estimation sample comprises these departments, where all agents have the same task and are evaluated based on the same performance measures. Handling inbound customer calls involves talking to customers as well as accessing and entering documentation in the customer database. Agents are not involved in any other task, such as written customer correspondence. Agents are assigned to team leaders whose main task is supervising the agents and monitoring their calls. Team leaders report to and are evaluated by the department managers.

6.2.2 Explicit incentives in the firm

Though agent performance is precisely measurable along several dimensions for any given time interval, explicitly formulated performance incentives to increase performance were only introduced in April 2011.⁴ In this new performance pay system agents are paid a wage premium if they outperform pre-defined performance thresholds on one of the observable performance outcomes.

Before the introduction of the performance bonuses evaluated in this chapter, the firm incentivised agents by an annual wage increase and an annual bonus. Agents were formally evaluated by their supervisor, i.e. their team leader, in April or May and would receive a grade from one (lowest) to five (highest).⁵ This grade was then

²In contrast to the other chapters of this thesis, this chapter uses a more recent sample period that covers the introduction of performance bonuses. In addition, this chapter uses information on more departments than the chapters of this thesis.

³There are also workers who receive bonuses based on their department's average performance because individual performance is not properly measurable, for example, general management and back-office workers.

⁴According to management, the main reason for the non-utilisation of explicit incentives based on observable performance outcomes was the position of the workers' council.

⁵Additional 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. Team leaders as well as

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%. There is no additional seniority-related wage increase employed at the call centre.

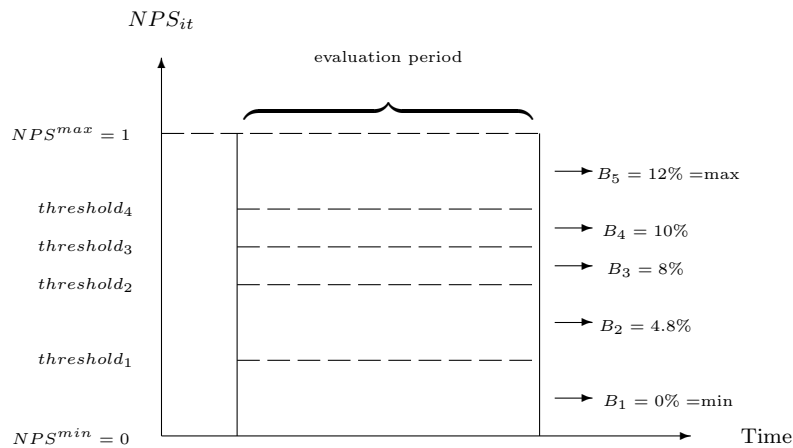
Though the annual performance rating by the supervisor had considerable impact on agents' wage growth and annual bonuses, the rating was at the discretion of the team leader and not (directly) dependent on an agent's observable performance outcomes. The team leader was supposed to consider several measurable performance outcomes, such as average handling time and measures of customer satisfaction, as well as the agent's behaviour towards peers, team leaders, and managers in the previous year. In addition to the lack of well-defined evaluation criteria such as performance thresholds, there was no pre-defined weighting of the different measures.

The firm decided to introduce a bonus that is explicitly related to service quality is that the quality of services provided to customers has become a unique competitive advantage in the mobile communications market. This new incentive scheme introduced in April 2011 only affected the bonus component. While annual performance rating was still used to determine an agent's annual wage growth, the firm abolished the annual performance bonus linked to the annual performance ratings. Under the new incentive scheme, up to 12% of an agent's wage is now paid as a bonus, depending on whether the agent outperforms the thresholds defined by management. The main difference is that the new incentive system is more precisely formulated in terms of specific service quality targets and is more transparent in terms of outcomes, which are clearly communicated before the bonus period. In addition, the evaluation periods are much shorter (three months) than that of the old incentive scheme (12 months), which increases worker awareness of the relation between their own performance and pay.

The new bonus is based on a measure of service quality gathered from customer satisfaction surveys of randomly phoned back customers. There are $j = 1, \dots, 5$ bonus levels B_j , which correspond to bonuses of 0, 4.8, 8, 10, and 12% of the wage earned, respectively. Agent performance is calculated and evaluated quarterly (Figure 6.1). If an agent's average performance \bar{y} does not exceed the lowest threshold y_1 ($\bar{y} \leq y_1$), the agent receives no bonus ($B_1 = 0\%$). Agents who outperform the highest threshold ($y_{J-1} \leq \bar{y}$) receive the highest bonus, B_J . In accordance with the

department managers were asked to reach a bell-curve distribution of performance gradings in both their teams and departments, respectively.

Figure 6.1: Design of the bonus system



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

bonus payments, the performance thresholds on which the bonuses are based are not equally distributed. It should be noted that the distance between the various thresholds differs. 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 are only 0.025.⁶ Throughout the quarter, agents are given feedback about their individual performance about once a week.

Agents are shown to experience substantial learning effects (for other performance outcomes, see Chapter 3 of this thesis). For this reason, management decided not to consider agents hired as temporary help or agents with less than six months' tenure for individual performance pay. Instead, these agents were paid a bonus depending on the average department performance.⁷

The calculated bonus pay is paid with the regular monthly wage 1.5 month after the end of each quarter. In the meantime, management announces the thresholds for the next quarter and communicates these to the team leaders and agents.

⁶Average performance differs by department. The target size is therefore adjusted accordingly. The absolute distance between the target thresholds is the same for each department.

⁷Furthermore, for agents whose quarterly performance is based on fewer than 60 evaluations, the agent's department average performance is used to assess their bonus level. In the estimation sample, about 45% of the agents did not have the minimum number of 60 evaluations during the evaluation period. These agents are more likely to be temporary workers with lower tenure, and workers who work less hours per week. Because agents are not able to affect whether customers participate in the survey, this should not bias the results.

6.2.3 Performance outcomes

While it is often difficult to directly measure worker performance, call centres have several measures covering different performance dimensions. These performance data comprise measures of service quality as well as measures that describe the work speed of agent tasks ('quantity'). 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. The availability of different performance dimensions, one of which is incentivised under the new incentive scheme, allows one to estimate the effect on both incentivised and non-incentivised measures. All performance measures are available at the individual agent level before and after the introduction of performance pay for a large number of time periods.

While several studies use team manager evaluations to measure worker performance, this information is potentially biased because of its subjectivity, and the reasons for this bias may be unobservable to the researcher (Flabbi and Ichino, 2001). All performance outcomes used in this chapter, however, were automatically generated and thus less prone to potentially subjective evaluations, as in the case of performance evaluations by team leaders or managers.

Service quality provided by agents

The newly introduced bonus pay in this call centre is based on a measure generated from a customer satisfaction survey among a randomly chosen population of customers. Amongst other questions, customers were asked whether they would recommend the mobile operator to family and friends, based on the previous call. This question could be answered on a scale from zero ('very unlikely') to 10 ('very likely').⁸

From the answers to this question, the management calculated the net-promoter score (see, e.g., Keiningham *et al.*, 2007), which is assumed to be related to customer loyalty. Because the question explicitly asks for the customer's opinion based on this call, I use this as a measure of an agent's service quality. This measure is defined as the percentage point difference between the share of customers rating the agent

⁸The exact question was 'Based on this contact, how likely are you to recommend [the firm] to your family and your friends?' The survey contained further information on whether the customer had already approached the call centre with the same problem previously, whether the problem that was the reason for the call had been solved, and how much effort the agent put into the call to solve the problem.

6 The heterogeneous effects of bonus pay on performance outcomes

as nine or 10 (high service quality) and those rating the agent as six or lower (low service quality):

$$NPS_{it} = y_{it}^q = \frac{N_{it,9-10} - N_{it,0-6}}{N_{it,0-10}} \quad (6.1)$$

where N is the number of evaluations and the subscript denotes the grade given by the customer for agent i in week t . For interpretation purposes, the customer satisfaction index used throughout this chapter is scaled between zero and one. All evaluations during the bonus period (i.e. the quarter) are used to calculate y_{it}^q and to assess the size of the bonus.

Before interacting with an agent, customers were asked whether they were willing to participate in a customer satisfaction survey. Importantly, the agent did not know whether the customer agreed to participate in the survey or not. Neither agents nor managers can affect the selection of customers who rate agent performance and thus cannot influence this outcome measure by selection into the survey. Shortly after the end of the customer call, a random subset of customers is automatically called back. An interactive voice response system then guides the customers through the survey.

Because service quality is taken from the customer satisfaction survey, the number of evaluations relative to the actual number of calls made is rather low. In the sample employed for this chapter, about 1.9% of the calls were evaluated by the customer satisfaction survey. This has two implications for the use of service quality as a performance measure: First, available agent–week observations of service quality are often based on only a low number of actual evaluations and thus have considerably more variation than the average. For this reason, all estimations control for the number of customer evaluations (see Section 6.3.3). Second, only about 61% of agent–week observations have at least one evaluation which substantially reduces the sample size for estimations using service quality as an outcome variable.⁹

Alternative performance outcomes

Apart from service quality y_{it}^q , which is used to calculate the bonus paid to agents, the call centre generates other performance outcomes that are not related to the new

⁹A potential concern about using y_{it}^q as a performance outcome is that it may be potentially biased because of customer non-response. Agents providing low-quality (high-quality) service would then be characterised by service quality that is higher (lower) than their actual service quality because the sample of evaluated calls is less representative of their calls than for agents providing higher (lower) quality. While this may be a potential concern about the validity of service quality as a proxy for quality provided, there is no reason to assume that customer non-response changed with the introduction of performance pay.

incentive scheme. These enable us to analyse the effect of incentives on performance outcomes that are not subject to the bonus pay calculation. This chapter uses performance outcomes that describe (1) work speed and (2) another service quality indicator that indicates whether the customer's problem was actually solved.

The performance outcome that describes work speed is based on the average length of calls to measure performance. Average handling time aht_{it} provides a clear and objective measure of quantitative performance that is available for each agent i and all calendar weeks 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. We therefore define the measure of (quantitative) performance as $y_{it} = \frac{100}{aht_{it}}$. Shorter calls with a lower average handling time aht_{it} are thus interpreted as higher performance y_{it} .

The share of problems solved is used as a second performance outcome to approximate service quality. The information is taken from the same survey as the incentivised performance outcome service quality. Customers were asked whether their 'question was completely resolved to' by the call agent in the corresponding call (yes/no). This variable is defined as the number of solved problems over all evaluations of agent i in week t . Compared to performance outcome service quality, the share of problems solved indicates whether the agent was able to solve the problem for which the customer called the call centre.

6.3 The effects of incentives on performance

6.3.1 Estimation sample

The data used in this chapter provide weekly information on the performance outcomes of the same workers before and after the introduction of the bonus related to service quality. We use the performance information on all agents who worked at least one week during the three months before or after the introduction of the performance bonuses. The sample also includes agents who were working in these departments but who were not (yet) eligible for individual performance pay because, for example, they did not meet the tenure or employment criteria.

The total number of agents in the estimation sample is 428, with 15,584 observations. The agents were observed on a weekly basis from week 6/2010 until week

Table 6.1: Descriptive statistics

	(1)	(2)
	Full sample	Estimation sample
Gender	0.408	0.404
<i>(share of male agents)</i>	(0.492)	(0.492)
Tenure	2.091	2.614
<i>(in years)</i>	(3.408)	(3.753)
Temporary help agent	0.612	0.589
	(0.430)	(0.430)
Age	30.085	31.183
	(9.611)	(10.118)
Average working hours	16.982	17.721
	(7.039)	(6.574)
Service quality	0.403	0.403
	(0.082)	(0.082)
Average handling time	0.326	0.329
<i>($y_{it} = \frac{100}{ah_{it}}$)</i>	(0.095)	(0.084)
Share of problems solved	0.698	0.699
	(0.111)	(0.110)
Number of evaluated calls	4.107	5.634
	(4.049)	(3.725)
Number of agents	429	312

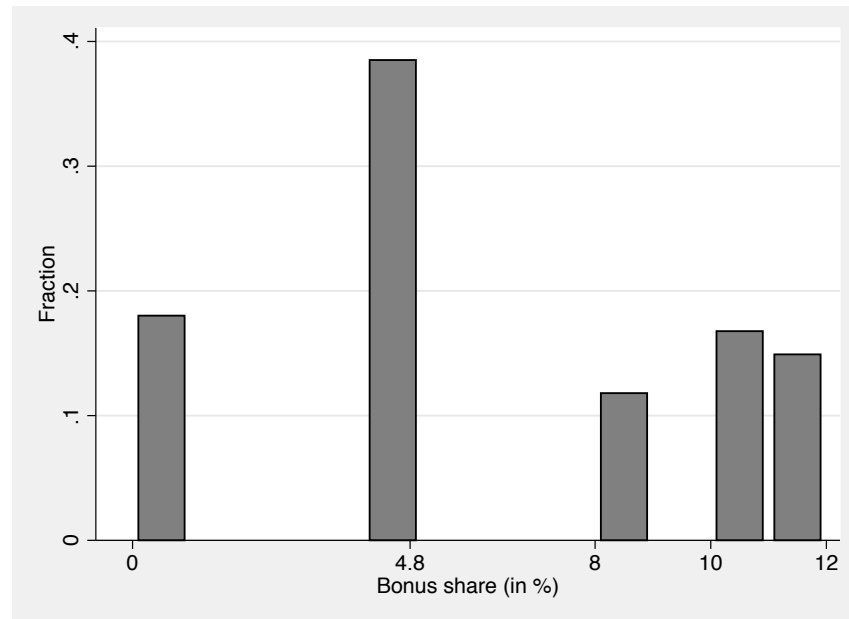
Standard deviation in parentheses.

26/2011 (end of quarter 2/2011).¹⁰ The sample thus consists of weekly data over more than four quarters before the introduction of performance pay and one quarter after. In the estimation sample, 41% of the agents are men; agents are, on average, 30 years old and have an average tenure of 2.1 years (Column (1) of Table 6.1). Across the estimation sample, 61% of all agent-week observations were working or had worked as temporary help agents who were not entitled to the bonus. Agents did, however, switch from temporary help agent contracts to fixed-term contracts within the sample period.

The average service quality provided by agents is 0.403, with a standard deviation of 0.082. This corresponds to an original value of about -0.25 on a scale from -1 to one. This score can be interpreted such that agents provide, on average, 25 percentage points fewer calls of high quality ($N_{it,9-10}$) than of low quality ($N_{it,0-6}$).

¹⁰Despite the fact that the same data are available for the time period after quarter 2/2011, they are not used in this chapter. This is because management re-evaluates the performance thresholds for each bonus quarter, which introduces endogeneity that can bias the estimation results. Furthermore, it is more difficult to establish causality because time trends hinder identification in the long run.

Figure 6.2: Distribution of bonuses paid in the first performance pay quarter



Note: This histogram shows the distribution of bonuses paid in the first quarter after the introduction of performance bonuses for service quality.

The mean work speed is 0.326, which corresponds to an average handling time of 5.11 minutes. On average, customers stated that agents solved more than two-thirds (0.698) of the problems for which they had contacted the call centre.

Column (2) of Table 6.1 shows the descriptive statistics for the sample with agent-week observations containing information on service quality. Of the 429 agents in the overall sample, 312 (73%) have information on service quality. A comparison of the means, however, shows that there is hardly a difference between agent-week observations including information on service quality and those that do not. There are, however, two exceptions: first, the average tenure is shorter in the full sample (2.09 years versus 2.61 years). The reason for this is that agents who leave the call centre early (short tenure) are undersampled in the estimation sample. Second, the average number of hours is higher for agents in the estimation sample (17.7 versus 17.0): Agents with shorter hours are more likely *not* to have an evaluation and are thus more likely not to have an *NPS*-observation in a week.

6.3.2 Descriptive results

Figure 6.2 shows the distribution of performance bonuses paid for the first quarter after the introduction of performance pay. On average, the bonus is 5.9% of the quarterly wage.

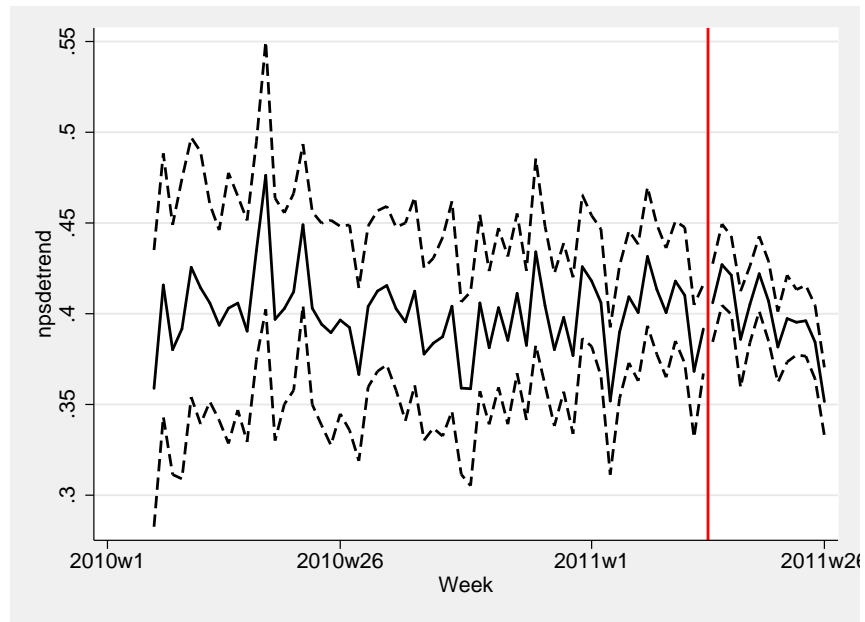
Table 6.2 shows descriptive statistics of the estimation sample, comparing the quarters before and after the introduction of the bonus. The table shows that the gender and tenure composition of the agents do not significantly differ between the periods before and after the introduction of the bonus. Furthermore, the table shows some evidence of the effect of the bonus introduction as service quality is slightly higher after the introduction though the difference is not significantly different from zero. The same holds for the two alternative outcomes of performance, share of problems solved and average handling time.

Table 6.2: Unconditional differences before and after the introduction of performance pay

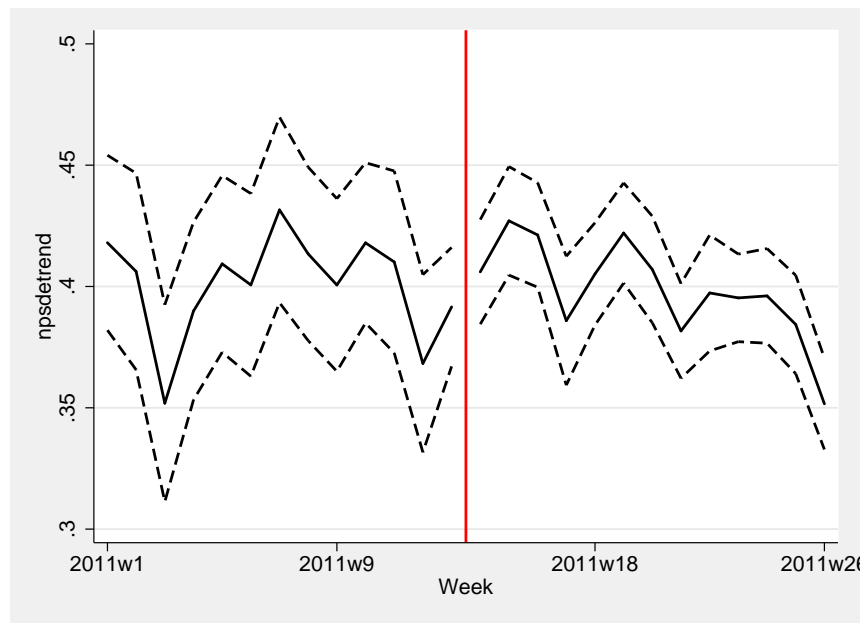
	(1)	(2)	(3)
Agents:	Fixed pay	Performance pay	Difference (2)-(1)
Gender	0.398	0.410	0.011
<i>(share of male agents)</i>	(0.026)	(0.026)	(0.036)
Tenure	2.670	2.584	-0.086
<i>(in years)</i>	(0.195)	(0.191)	(0.273)
Temporary help agent	0.512	0.469	-0.043
	(0.025)	(0.025)	(0.036)
Age	30.738	30.367	-0.371
	(0.544)	(0.530)	(0.761)
Working hours	16.108	16.141	0.033
	(0.439)	(0.414)	(0.603)
Service quality	0.407	0.416	0.010
	(0.008)	(0.006)	(0.008)
Average handling time	0.320	0.327	0.007
<i>($y_{it} = \frac{100}{ah_{it}}$)</i>	(0.006)	(0.005)	(0.008)
Share of problems solved	0.711	0.714	0.004
	(0.010)	(0.008)	(0.013)
Number of evaluated calls	2.528	8.067	5.538
	(0.129)	(0.329)	(0.358)
Number of agents	359	371	

Standard errors in parentheses. Descriptive statistics are averaged for the quarter before and the quarter after the introduction of performance pay.

Figure 6.3: Service quality before and after the introduction of performance bonuses



(a) Whole sample period



(b) Quarters before and after the introduction

Note: This figure shows the average levels of service quality y_{it}^q before and after the introduction of incentives (vertical line). Observations are weighted by the number of evaluated calls and detrended in a regression with a linear and a squared time trend included.

Panels (a) and (b) of Figure 6.3 depict the development of service quality over time. Despite the fact that service quality varies over time, it does not seem to

follow seasonal patterns. In line with the descriptive results shown in Table 6.2, this figure suggests no significant increase in performance, despite the fact that time effects are controlled for by including linear and quadratic time trends. The figure shows, however, that the line seems to become less volatile after the introduction of performance pay.

6.3.3 Estimation strategy

The effect of the introduction of explicit performance bonuses on the incentivised performance outcome is estimated by a regression of the output y_{it}^q on a dummy variable that equals one in the period after the introduction and zero before (p_{it}):

$$y_{it}^q = \alpha_i + \tau p_{it} + \beta_1 t_t + \beta_2 X_{it} + u_{it} \quad (6.2)$$

where the α_i are individual fixed effects to account for individual heterogeneity, t_t contains controls for overall time trends, the X_{it} are covariates such as working hours in week t and agent tenure, and u_{it} is an idiosyncratic error term.¹¹ Throughout all regressions, standard errors are clustered at the agent level.¹²

The identification strategy to estimate the effect of bonus pay on the underlying performance outcome relies on the comparison of the performance outcomes of the (same) workers before and after the introduction of the bonus pay. This requires appropriately controlling for other, potentially confounding effects such as trends in performance and seasonal effects.¹³

Because appropriately controlling for time trends is important in this setting, where *all* agents are subject to the new pay scheme, the regressions include linear and squared time trends. In addition, dummies for the month of the year are used to take seasonal patterns into account.

¹¹Throughout this chapter, the regressions do not control for individual tenure because tenure, individual fixed effects, and the linear time trend are perfectly collinear.

¹²The regressions using service quality and share of problems solved as measures of outcome are weighted by the number of evaluated calls in week t . This weighting is introduced because this weekly number of evaluated calls varies, which potentially affects the accuracy of the measurement of the outcome variable. In addition, these estimations contain dummy variables for each possible number of evaluations to control for within-agent differences in performance outcomes due to the varying number of evaluations.

¹³An alternative option to identify the effect of performance bonuses on individual performance would be to use ineligible workers, i.e. temporary help agents and low-tenure agents, as a control group in a difference-in-difference setting. Because these agents are mostly newly hired and show strong, non-linear growth rates in performance (see Chapter 3), this type of analysis may easily underestimate the true treatment effect.

Furthermore, estimation of the effect of bonus pay on performance outcomes may be confounded by other, for example, organisational changes within the firm that took place at the same time as the introduction of the bonus pay. In January 2011, 2.5 months before the introduction of the bonus pay, the departments were reorganised. However, no other organisational changes took place at the same time as the performance bonuses were introduced, or any other changes, such as sudden changes in the structure of calls, for example, due to newly introduced products. The estimated effect of the treatment dummy p_{it} ($\hat{\tau}$) should thus provide an unbiased estimate of the effect of the introduction of the bonus related to service quality.

6.3.4 Estimation results

Average effects

Using information on performance outcomes at the individual level before and after the introduction of the individual performance bonuses, Table 6.3 shows the results when estimating Equation (6.2) with service quality as the dependent variable. As a baseline estimate, Column (1) of Table 6.3 shows the effect of the introduction of performance pay on service quality controlling only for a linear and a quadratic time trend. Conditional on time trends, performance is 0.027 higher after the introduction, of individual performance bonuses, compared to before. This corresponds to about a third of the standard deviation in service quality.¹⁴

In Columns (2) and (3), I further control for the department the agent is working for, the agent's number of working hours in a week, whether the agent is employed as a temporary help agent (and therefore is not eligible for the bonus pay), and agent-level individual fixed effects. The estimated effect slightly decreases to 0.022 and 0.023 in Columns (2) and (3), respectively, which corresponds to 28% of one standard deviation of service quality.

These results show that workers do react to the incentives set by management by providing higher service quality to customers.

Alternative performance outcomes

Though the bonus is based on service quality only, one can expect an effect on non-incentivised performance outcomes as well. The first argument for an effect on non-incentivised outcomes is that agents will focus on the incentivised outcome,

¹⁴Estimation results excluding agents who were not eligible due to tenure or their working contract shows similar results as the ones presented in this section.

Table 6.3: The effect of performance pay on service quality

	(1)	(2)	(3)
Bonus pay dummy	0.027*** (0.007)	0.022*** (0.007)	0.023*** (0.007)
Time trend	0.002*** (0.001)	0.003*** (0.001)	0.003*** (0.001)
Time trend ² /10	-0.000** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)
Working hours		-0.001*** (0.000)	-0.001** (0.000)
Temporary help agent		-0.007 (0.005)	-0.007 (0.008)
Constant	0.357*** (0.015)	0.338*** (0.020)	0.331*** (0.023)
Controls (department)	No	Yes	Yes
Individual fixed effects	No	No	Yes
R^2	0.012	0.058	0.140
N (agents)	312	312	312
N	9525	9525	9525

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: y_{it}^q . All regressions are weighted by the number of evaluated calls in week t for agent i and include dummies for the number of evaluated calls for each agent-week observation. All standard errors are clustered at the agent level.

thereby ‘neglecting’ other, potentially conflicting performance outcomes. If performance outcomes conflict one expects a negative effect on other performance outcomes and a positive effect if these are positively related to each other. The second argument for externalities on other performance outcomes is that these were implicitly incentivised under the old incentive scheme, where *any* behaviour mattered for the annual performance rating and thus for the annual bonus paid.

Figure 6.4 shows the development before and after the introduction of performance pay for average handling time (Panel (a)) and the share of problems solved (Panel (b)). Over the sample period, average handling time shows a downward trend before the introduction.¹⁵ The overall correlation between service quality and average handling time *before* the introduction is -0.039 (significant at the 5% level). On the other hand, the share of problems solved, which has a strong positive correlation with service quality (0.38), shows a relatively stable pattern before and after the introduction of performance pay.

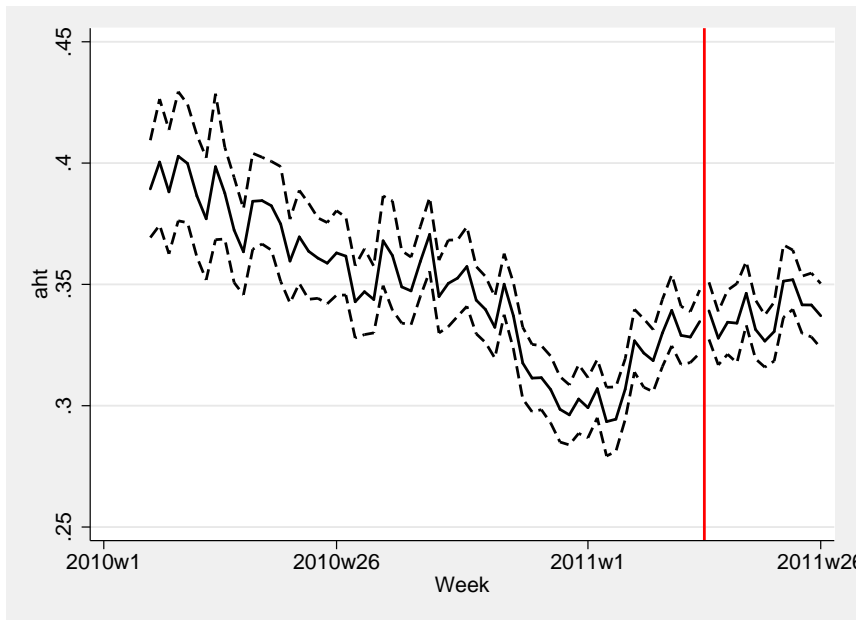
To test whether the introduction of incentivising service quality affects alternative performance outcomes, the regressions shown in Table 6.3 are replicated with the performance outcomes’ average handling time and share of problems solved. For average handling time (Columns (1) to (3) of Table 6.4), the results show a significant increase of 0.021 to 0.027, respectively, which corresponds to an increase of roughly one-fourth of a standard deviation in average handling time. For the share of problems solved, the effect is slightly higher, ranging from 0.021 to 0.029 (one-fourth of one standard deviation).¹⁶

¹⁵According to the firm’s management, this was mainly due to the fact that agents were pushed to perform better on service quality than on other measures, such as average handling time.

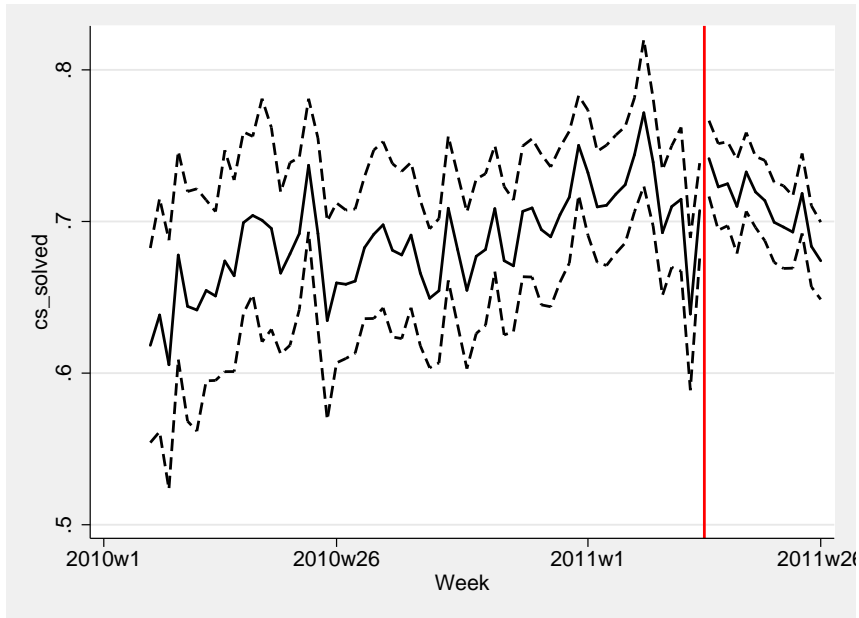
¹⁶The number of observations is smaller in regressions where the share of problems solved is the measure of outcome because this measure is based on the aggregate of customer evaluations of agent i in week t .

6 The heterogeneous effects of bonus pay on performance outcomes

Figure 6.4: (a) Average handling times and (b) shares of problem solved before and after the introduction of performance bonuses



(a) Average handling time



(b) Share of problems solved

Note: Panel (a) shows the average handling performance outcomes before and after the introduction of incentives (vertical line); Panel (b) shows the corresponding outcomes for the share of problems solved. The observations in Panel (b) are weighted by the number of evaluated calls.

Table 6.4: The effect of the introduction of performance bonuses on average handling time and share of problems solved

	(1)	(2)	(3)	(4)	(5)	(6)
Performance outcome	Average handling time			Share of problems solved		
Bonus pay dummy	0.021*** (0.005)	0.027*** (0.005)	0.021*** (0.005)	0.021** (0.009)	0.028*** (0.009)	0.029*** (0.009)
Time trend	-0.003*** (0.001)	-0.001** (0.001)	-0.001** (0.001)	0.004*** (0.001)	0.005*** (0.001)	0.004*** (0.001)
Time trend ² /10	0.000*** (0.000)	-0.000 (0.000)	0.000 (0.000)	-0.000*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Working hours		-0.002*** (0.000)	-0.002*** (0.000)		-0.001*** (0.000)	-0.001** (0.000)
Temporary help agent		(0.000) -0.023*** (0.006)	(0.000) 0.001 (0.006)		(0.000) -0.000 (0.006)	(0.000) -0.005 (0.010)
Constant	0.411*** (0.012)	0.429*** (0.016)	0.399*** (0.015)	0.618*** (0.018)	0.623*** (0.024)	0.638*** (0.025)
Weighting	No	No	No	Yes	Yes	Yes
Controls (department)	No	Yes	Yes	No	Yes	Yes
Individual fixed effects	No	No	Yes	No	No	Yes
R^2	0.038	0.239	0.553	0.018	0.078	0.165
N (agents)	428	428	428	312	312	312
N	15584	15584	15584	9495	9495	9495

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: average handling time (Columns (1) through (3)); share of problems solved (Columns (4) through (6)). Regression presented in Columns (4) through (6) furthermore include dummies for the number of evaluated calls for each agent-week observation. All standard errors are clustered at the agent level.

These findings suggest that the introduction of bonus pay based on incentivised performance outcomes does not have any negative effects on non-incentivised performance outcomes. The results for the average handling time could be explained by the fact that they are relevant to the annual wage increase. However, this result may also be explained by the argument that customers simply prefer shorter calls, irrespective of the outcome of the call. The effect on the share of problems solved is likely due to the complementary relation with service quality.

These results provide evidence against the multitasking hypothesis (Holstrom and Milgrom, 1991) because agents do perform better, even on non-incentivised performance outcomes. Despite the fact that the relative importance of different performance outcomes has changed (i.e., service quality is incentivised with the introduction of the bonus pay), the alternative performance outcomes may still matter for the annual wage increase or more generally for promotion decisions.

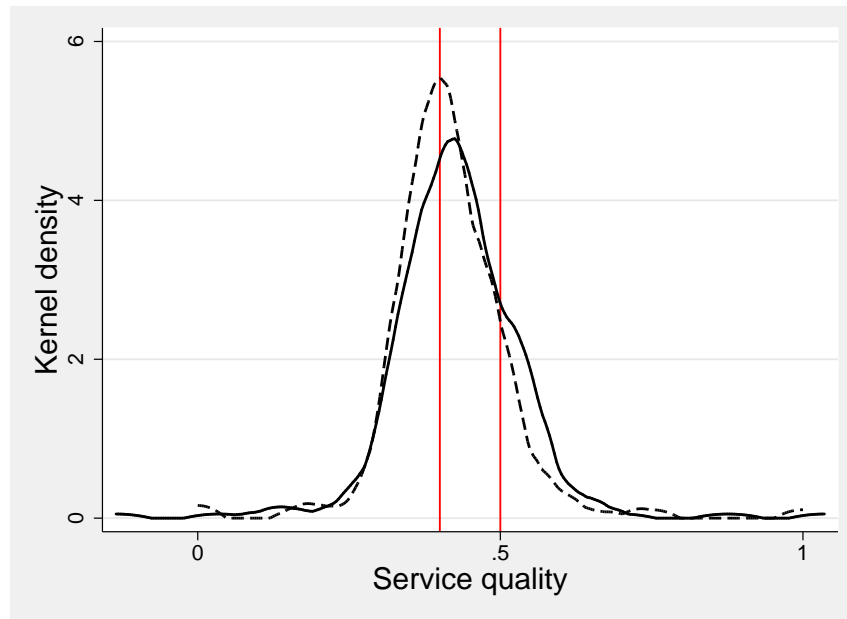
Heterogeneous effects

Under performance pay, agents receive individual performance bonuses when outperforming performance thresholds on service quality that are set in advance by management. The design of the incentive scheme implies that agents of high ability, who easily outperform the highest performance threshold, do not have an incentive to further improve their performance since they will receive the maximum bonus in any case. On the other hand, low-ability agents, who perform below the lowest threshold, do have an incentive to increase their efforts to achieve a bonus by reaching (one of) the performance thresholds (Figure 6.1).

The descriptive evidence suggests that there are indeed heterogeneous effects due to the design of the incentive scheme applied by the firm. Figure 6.5 shows the kernel distribution of agents' average service quality in the quarters before (solid line) and after (dashed line) the introduction of performance pay. The vertical lines show the lowest and highest performance thresholds, respectively. Below the lowest threshold, agents receive no bonus, while agents above the highest threshold receive the maximum bonus of 12% of their wage. The figure suggests that the distribution in the lower part moves slightly to the right (an increase in performance). However, the performance to the right of the highest performance threshold moves slightly to the left (a decrease in performance). This suggests that the effect of the bonus on the performance of the high-ability agents is actually negative.

This pattern is confirmed by the estimation results shown in Panel (a) of Table 6.5. The table replicates the regression used in Column (3) of Table 6.3 by ability quartile.

Figure 6.5: Distribution of service quality by payment scheme



Note: This figure shows the distributions of service quality y_{it}^q in the last quarter under fixed pay (solid line) and in the first quarter under performance pay (dashed line). Vertical lines denote the lowest and highest performance thresholds for achieving the bonus under the new incentive scheme.

For this purpose, the sample is differentiated by the average performance in the quarter *before* the introduction of the performance pay. In line with predictions, the results show that the effect is highest for the lowest quartile of the ability distribution (0.088) and negative (−0.034) for the highest quartile (high-ability agents). The effect for the second quartile is roughly the same as the average effect, whereas performance of the third quartile did not increase significantly with the introduction of bonus pay. While the average effects are positive (Table 6.3), these results show that this effect is driven by a strong increase for low-ability agents and even a significant decrease for high-ability agents.

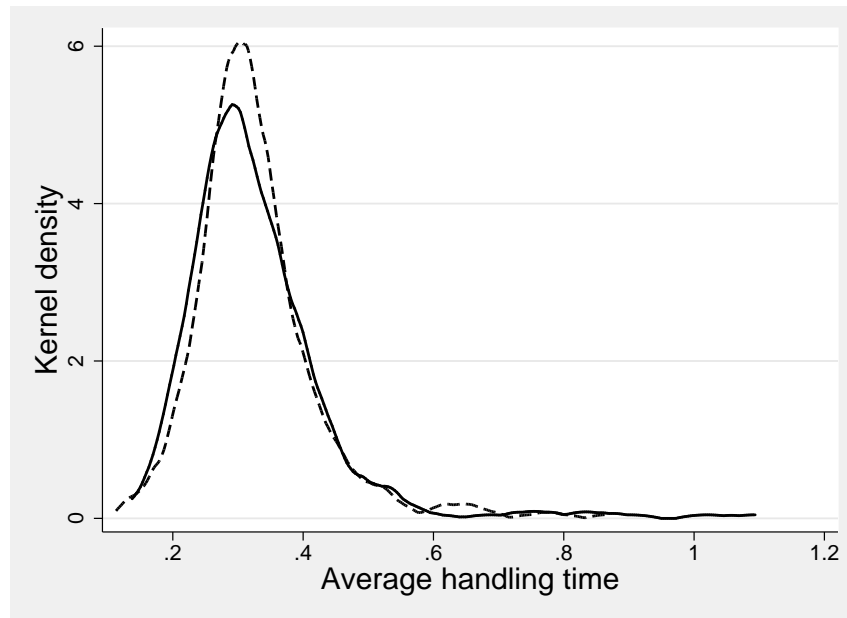
For the two alternative performance outcomes not subject to the new bonus pay, the results show slightly different patterns. Panel (a) of Figure 6.6 shows the kernel density of the average handling time in the quarters before and after the introduction of performance pay. It suggests that there is only a moderate increase in average handling time related to the introduction of the bonus pay scheme. In line with this figure, Panel (b) of Table 6.5 shows that the increase in average handling time is driven by the second to fourth quartiles (0.029–0.037), while the effect is insignificant for the lowest quartile.

Table 6.5: Heterogenous effects of the introduction of performance bonuses on performance outcomes

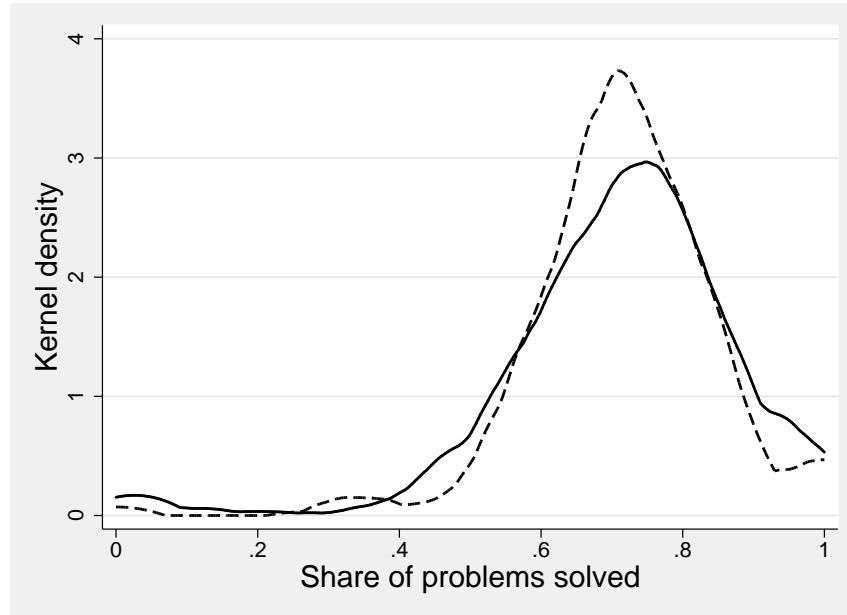
Ability quartile	(1)	(2)	(3)	(4)
(a) Dependent variable: Service quality				
Bonus pay dummy	0.088*** (0.014)	0.034*** (0.012)	-0.007 (0.012)	-0.034** (0.015)
R^2	0.125	0.114	0.106	0.161
N (agents)	65	62	58	63
N	2303	2595	2243	2063
(b) Dependent variable: Average handling time				
Bonus pay dummy	0.016 (0.012)	0.034*** (0.007)	0.037*** (0.008)	0.029** (0.011)
R^2	0.426	0.660	0.605	0.497
N (agents)	65	62	58	63
N	3112.000	3083.000	3083.000	3029.000
(c) Dependent variable: Share of problems solved				
Bonus pay dummy	0.088*** (0.021)	0.024 (0.016)	0.006 (0.018)	-0.005 (0.018)
R^2	0.191	0.139	0.124	0.180
N (agents)	65	62	58	63
N	2295	2587	2237	2056

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. All regressions for *service quality* and *share of problems solved* are weighted by the number of evaluated calls in week t for agent i and include dummies for the number of evaluated calls for each agent-week observation. All regressions control for department, and individual fixed effects. All standard errors are clustered at the agent level. Ability quartiles are generated based on agents' average service quality in the quarter *before* the introduction.

Figure 6.6: Distribution of (a) average handling times and (b) share of problem solved by payment scheme



(a) Average handling time



(b) Share of problems solved

Note: This figure shows (a) the distribution of average handling times and (b) the share of problems solved in the last quarter under fixed pay (solid line) and in the first quarter under performance pay (dashed line).

Table 6.6: Seasonal effects and performance pay

Bonus pay dummy	0.050** (0.021)
Time trend	0.006*** (0.001)
Time trend ² /10	-0.001*** (0.000)
Constant	0.243*** (0.036)
Seasonal controls	Yes
R^2	0.143
N (agents)	312
N	9525

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: y_{it}^q . Regressions are weighted by the number of evaluated calls in week t for agent i . All standard errors are clustered at the agent level. All regressions include controls for individual fixed effects, department dummies, controls for working hours, employment status, and dummies for the number of evaluated calls for each agent-week observation. Seasonal controls include dummies for the month of the year.

The effect for the share of problems solved is similar to the effect of service quality: Panel (b) of Figure 6.6 suggests a small increase in the lower part of the distribution. Indeed, the estimation results in Panel (c) of Table 6.5 show that the effect is significantly positive only for the lowest quartile while it is insignificant for the other quartiles. As for service quality, this result shows that the effect is driven by low-ability agents for whom the incentive to perform better on service quality is higher than for high-ability agents.

6.4 Robustness analyses

6.4.1 Seasonality and confounding factors

Compared to settings where the researcher decides the timing of the treatment or observes randomised treatment and control groups at the same time (see Bandiera *et al.*, 2007; Shi, 2010; and Chapter 4), the bonus pay was introduced for all agents at the same time. The effects are thus identified by observing the same workers over time, before and after the introduction of bonus pay. A major concern about the identification strategy is that general time effects, such as seasonal patterns, may confound the estimated treatment effect. If, for instance, the treatment (post-introduction) period is characterised by higher performance in each year, seasonal effects may confound the estimated treatment effect.

Table 6.7: Sample length and the estimated treatment effect

	(1)	(2)	(3)	(4)	(5)
Sample period	2010w14-2011w26	2010w27-2011w26	2010w40-2011w26	2011w1-2011w26	2011w10-2011w17
Bonus pay dummy	0.023*** (0.007)	0.027*** (0.007)	0.026*** (0.008)	0.018** (0.009)	0.032** (0.016)
Time trend	0.004*** (0.001)	0.007*** (0.001)	0.011*** (0.003)	0.025*** (0.008)	-0.045 (0.116)
Time trend ² /10	-0.000*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.002*** (0.001)	0.003 (0.009)
Constant	0.311*** (0.031)	0.236*** (0.050)	0.143 (0.096)	-0.286 (0.246)	1.913 (3.565)
R^2	0.145	0.154	0.170	0.206	0.254
N (agents)	311	311	310	309	247
N	9001	7995	6776	5155	1605

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: y_{it}^g . Weighted regressions are weighted by the number of evaluated calls in week t for agent i . All standard errors are clustered at the agent level. All regressions include controls for individual fixed effects, department dummies, controls for working hours, employment status, and dummies for the number of evaluated calls for each agent-week observation.

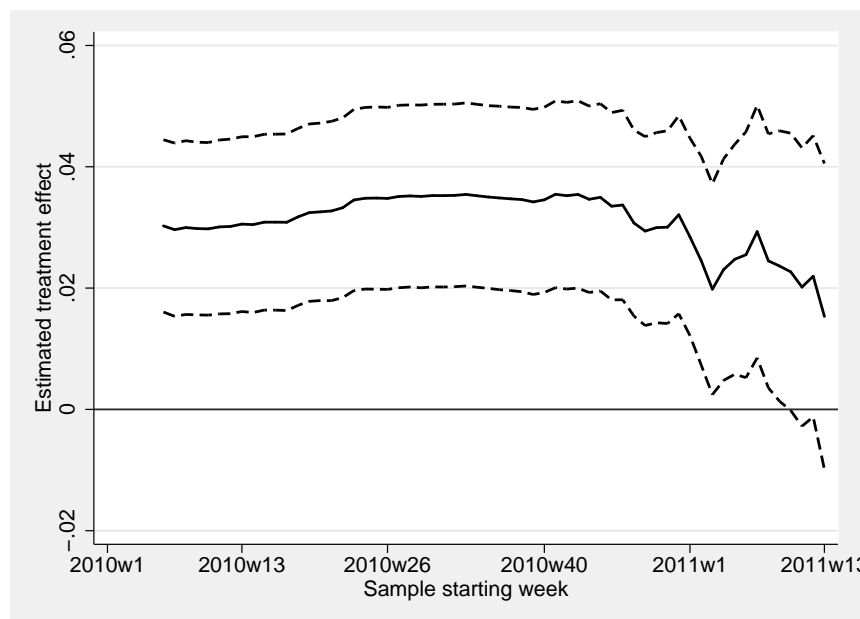
To explore whether the estimated treatment effect is contaminated by seasonal effects, the main equation (shown in Column (4) of Table 6.3) is estimated in different specifications and for different sample periods around the introduction date of the performance pay. As first part of the evidence, Table 6.6 shows the main regression, with month of the year dummies added. These variables should pick up recurring seasonal variation that is not captured by the time trends included. The estimated treatment effect is slightly higher than for the regression shown in Table 6.3.

As a second piece of evidence against the alternative hypothesis of time effects, Table 6.7 show the main regression, for varying time periods. First, Columns (1) through (3) show the estimated treatment effects if the sample period starts later than in the main results shown in the previous section. For this purpose, Column (1) shows the results if the sample starts in the second quarter of 2010, Column (2) for the sample starting in the third quarter of 2010, and Column (3) for the sample starting in the fourth quarter of 2010. Irrespective of the length of the sample, these results show that the estimated treatment effect $\hat{\tau}$ remains stable. This result can also be visualised in Figure 6.7, which shows the estimated treatment effect for sample periods that start later than that used in this chapter. The figure shows that the estimated treatment effect is relatively stable and significant.

Columns (4) and (5) show the estimated treatment effect when the estimation sample contains only the quarters before and after the introduction and just the months before and after, respectively. If there are no announcement effects, the estimated treatment effect should also be determined when comparing the periods just before and just after the introduction of the bonus pay. The results show that both the quarter and month samples are significant, with 0.018 and 0.032, and comparable to the main results.

One may also be concerned about the announcement effect of the introduction of the bonus pay. The agents were first informed about the fact that service quality would become the basis for bonus pay in November 2010. A reorganisation of the departments that came along with the introduction of performance pay was carried out in January 2011. In the second half of March 2011, agents were informed that they would receive a bonus on top of their wage paid from April 2011. To explore this, the baseline regression is estimated with hypothetical dates for the introduction, such as the date of the announcement instead of the actual introduction. If this leads to a significant treatment effect, such evidence would suggest that the announcement already affected worker performance. Figure 6.8 plots estimated treatment effects for any possible hypothetical date. The vertical lines show four important dates: (1)

Figure 6.7: The effect of shorter sample periods on the estimated treatment effect



Note: This figure shows the estimated treatment effect when the sample starts with the week mentioned on the x -axis (instead of 2010w6). The underlying regression is a regression of service quality on a linear and a quadratic time trend and worker fixed effects. The dashed lines show the corresponding 95% confidence interval.

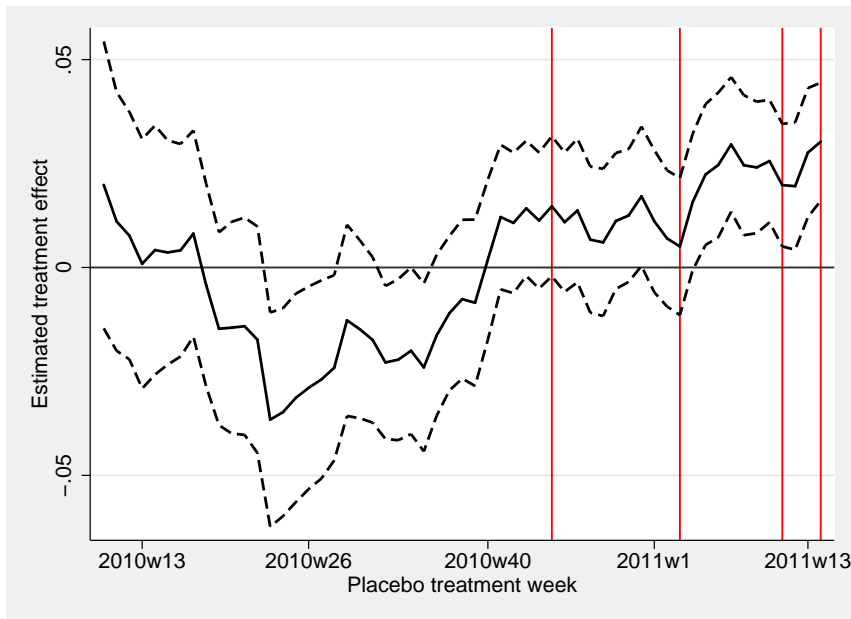
the announcement, (2) the reorganisation, (3) the actual announcement about the details and the starting date, and (4) the actual introduction.

The estimated treatment effects plotted in Figure 6.8 show that neither the announcement date (1) nor the reorganisation date (2) is associated with a significant treatment effect. When I define the treatment dummy as being one from any week between January to March 2011 onwards, however, the estimated treatment effects are significantly different from zero. This suggests that a part of the treatment effect can be explained by announcement effects, that is, that agents improved their performance already *before* the actual introduction of the bonus pay.¹⁷

Furthermore, Figure 6.8 shows evidence that the date of the introduction of the bonus scheme is not prone to (recurring) seasonal effects. The figure shows that when the introduction is artificially set to $t - 1$ (year), the estimated treatment effect is not significantly different from zero. Besides general time trends, however, the estimated treatment effect may also be the result of individual trends, for instance, due to learning on the job. Chapter 3 of this thesis shows that learning is most important in the first year of working for the call centre.

¹⁷This can also be tested by including a dummy for the period between announcement and actual introduction. In this case, the estimated treatment effect increases to 0.047*** with the pre-introduction dummy being 0.022 and significant at the 10% level.

Figure 6.8: The effect of placebo treatments on the estimated treatment effect



Note: This figure shows the estimated treatment effect when shifting the introduction hypothetically to the week mentioned on the x -axis (instead of the actual introduction in week 2011w14). The underlying regression is a regression of service quality on a linear and a quadratic time trend and worker fixed effects. The dashed lines show the corresponding 95% confidence interval.

To investigate this effect, Table 6.8 shows the baseline regression augmented by the interaction of the linear time trend and individual fixed effects. This takes any (heterogeneous) learning effects at the individual level into account. The estimation results show that the estimated treatment effect remains almost unchanged compared to the baseline results shown in Column (3) in Table 6.3.

6.4.2 Sorting and turnover

Previous studies show that changes in incentives also change the ability distribution in the workplace (Lazear, 2000). In call centres, turnover is traditionally relatively high (Batt *et al.*, 2005). Figure 6.9 shows the separation rates over time.¹⁸ Separation rates show a peak in the quarter before the introduction, suggesting that agents left the firm to avoid performance pay.

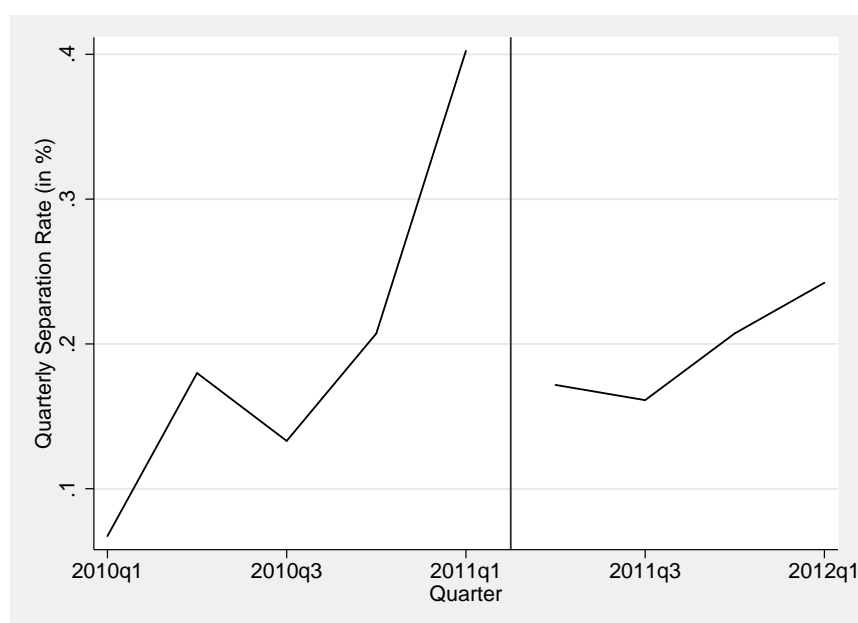
¹⁸For the purpose of analysing the effects of sorting, a longer sample is used that considers observation until the first quarter of 2012.

Table 6.8: The effect of performance pay on service quality and individual time trends

	(1)
Bonus pay dummy	0.024*** (0.007)
Time trend	-0.000 (0.000)
Time trend ² /10	-0.000*** (0.000)
Working hours	-0.000 (0.000)
Temporary help agent	0.005 (0.010)
Constant	0.393*** (0.026)
R^2	0.180
N (agents)	312
N	9525

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Dependent variable: y_{it}^q . The regression is weighted by the number of evaluated calls in week t for agent i and controls for department and individual fixed effects, working hours, employment status, and dummies for the number of evaluated calls for each agent-week observation. Furthermore, the regression contains all interactions between individual fixed effects and a linear time trend. All standard errors are clustered at the agent level.

Figure 6.9: Turnover rates over time



Note: This figure shows the turnover rate (ratio of the number of separations per quarter to the overall number of workers per quarter) by quarter.

To analyse sorting, I use a measure of ability that is calculated from the fixed effects coefficients in a regression controlling for tenure and other covariates. When comparing the ability of agents who left in the quarter before the introduction (average ability 0.313) with that of agents who left in the quarter after the introduction (0.361). The agents who left their job just after the introduction are of about 10 percentage points higher ability than the agents who left before the introduction.

The picture is different, when comparing the ability measure for new starters just before and just after the introduction of performance pay. This allows to analyse whether incoming agents differ from before to after the introduction of performance pay. The average ability for starters is 0.400 before the introduction of the performance bonus and 0.384 afterward. The difference is, however, not significant.

6.5 Conclusion

This chapter analyses the effect of the introduction of bonus pay on worker performance using unique data on agents working in the call centre in a multi-national telephone company. The data contain qualitative as well as quantitative performance information on performance before and after the introduction of a performance bonus. The bonus pay, however, was based merely on service quality performance outcomes, while other performance outcomes such as average handling times, were not incentivised.

We find that agents increase their service quality by about a third of a standard deviation. This shows that agents do react to the incentives set by the management and increase their performance to get a monetary bonus. Furthermore, I show that agents also perform better on alternative outcomes, namely work speed and share of problems solved. This result may be explained by the fact that alternative performance outcomes may still matter for less well-defined incentives, such as promotion decisions. It may also be explained by the complementarity of the performance measures: Customers may appreciate shorter calls and therefore rate agents providing shorter calls better.

Furthermore, the results show that the estimated effect of the bonus introduced is driven by low-ability agents. These agents react much stronger to the incentives set, while high-ability agents even reacted negatively to the incentives.

7 Conclusion

7.1 Summary

The aim of this thesis was to determine the role of formal and informal learning as well as explicit and implicit incentives for worker performance. For this purpose, I employed personnel data gathered in a call centre of a multi-national telecommunications firm. These data allow to follow a large number of workers with very detailed information on several performance outcomes over several years. In Chapter 2, I provide details on these rich data.

In Chapter 3, I analysed tenure-performance profiles of call agents, using these unique panel data that contain detailed performance information on the individual level. The results show that call agents have a very steep tenure-performance profile in the beginning of their job in the call centre. This performance increase, however, flattens considerably after about 6 to 8 months. The tenure-performance profile can very well be described by a logarithmic function. An increase of tenure by 10% leads to an increase of 5.5% of one standard deviation in performance. This increase in performance can be interpreted as accumulation of human capital due to learning by doing. This explains why the highest returns to learning occur in the beginning of the employment relationship. The data show that agents perform 66.5% better after 8 months compared to their initial performance. This huge increase in performance after workers have built up some experience in the job is comparable to related findings from other sectors of industry (see, e.g., Shaw and Lazear, 2008). After this initial period, the marginal return to tenure decreases. In addition, I show that workers significantly benefit from working with more experienced peers. I find that an increase in average team tenure of one standard deviation relates to an increase in performance of 14% of one standard deviation. This shows that it can be beneficial for firms to place new workers in teams with more experienced agents in order to facilitate knowledge spill-overs between workers. My findings show that this holds even in a case of a call centre team where workers are not involved in team work.

In Chapter 4, which analyses the effect of training participation on worker performance, I randomly assigned agents to training groups, thereby generating exogenous variation in training participation. Regression results show that the performance of agents improved by about 10% after following the training. This return is much higher than the returns to training found in studies that attempt to correct for selectivity. This may be explained by the fact that the effects of formal training on productivity are much greater than the returns measured by workers' wages (Dear- den *et al.*, 2006). However, I also expect that returns to training in the call centre sector are relatively high because the sector lacks proper vocational education for

their core jobs and therefore needs to develop the human capital required for a good performance in the job by training its workforce. Furthermore, I find that the performance increase after the training is not due to lower quality provided by the worker. Instead, training participation also has a positive effect on customer ratings on agents' knowledge and understanding of the customers' question. Moreover, I show that agents who did not participate in the training also improved their performance through externalities. I find that increasing the share of trained peers in the team by 10 percentage points leads to a 0.51% increase in performance of the agents who are not yet trained.

In Chapter 5, which examines implicit incentives, I analyse the effect of implicit incentives on worker's performance by analysing the extent to which team leaders can influence worker productivity by exerting pressure on their subordinates who are performing relatively poorly. Team leaders are expected to put pressure on call agents to increase their performance and thus achieve high average team performance. By ranking the call agents within teams, team leaders can influence their agents' effort provision. I show that the effects are stronger for high-performing workers than for low-performing workers, independent of the pressure from within-team rankings or team rank. Furthermore, I show that the effects of individual rankings are much larger than the effects of team rankings.

In Chapter 6, I analyse the effect of an explicit incentive: the introduction of bonus pay on worker performance. While the data contain performance outcomes that describe service quality ("net promoter score") as well as performance outcomes that describe work speed ("average handling time"), the bonus was merely based on the first performance measure, service quality. I find that, after the introduction of the bonus scheme, agents increase their service quality, on average, by about a third of a standard deviation. This shows that agents do react to the incentives set by the firm management and increase their performance to get the bonus. Furthermore, I show that agents also perform better on alternative outcomes, namely their work speed and share of problems solved. This result may be explained by the fact that alternative performance outcomes may still matter for less well-defined incentives, such as promotion decisions. In addition, the results show that the estimated effect of the bonus introduced is driven by low-ability agents. These agents reacted much more strongly to the explicit incentives set, while high-ability agents even reacted negatively to the incentives.

7.2 Contributions and implications for human resource management

The evidence on the effect of both learning and incentives (Chapters 3 through 6) shows how different types of HR practices affect workers' performance.

Chapters 3 and 4 focus on the role of learning for workers' performance. Chapter 4, which analyses the effects of formal training, is the first study to provide experimental evidence on the causal effects of investments in training on productivity using individual performance information. From a policy point of view, an unbiased estimation of the impact of training on productivity is important for assessing the role of further training in the development of human capital. Moreover, information on the causal effect of training on productivity is crucial for convincing firms' top management that high investments in training courses are justified. Most previous studies have shown that when appropriately taking selection bias into account, the estimated effects of training disappear. In Chapter 4, I show that formal training has a positive impact on performance, even with exogenous variation in training participation. This may, however, also be due to the fact that there is no initial vocational training for call agents. Therefore, skills required in the job have to be acquired through in-house formal training programmes. Call centres traditionally have relative high labour turnover inducing a high risk of losing their training investments. According to human capital theory, call centres would then invest less in low-return training programmes than firms in sectors with lower turnover rates.

Besides learning through formal training programmes, workers also acquire human capital through learning on the job. In Chapter 3, I show that call agents show substantial growth rates in performance over the first 8 months. This performance increase can be interpreted as human capital acquisition by means of learning by doing. Calculations show that the investments in informal learning are almost twice as large as firm's investments in formal training programmes in terms of hours. This suggests that firms should not only consider formal training programmes as important, but also consider learning on the job as an important source of human capital acquisition. However, this result also implies that labour turnover, which is traditionally very high in the call centre sector, is even more costly because large investments in human capital in the early career of call centre agents do not pay off in the long run. It is difficult to compare the degree of informal learning across jobs and across sectors. It is likely that tenure-performance profiles are flatter in more skill-intensive jobs and even steeper in jobs with lower skill requirements.

7.2 Contributions and implications for human resource management

I also analysed the role of peer effects in human capital acquisition. In both Chapters 3 and 4, I show that working with well-performing and trained peers, respectively, significantly increases worker performance. These results suggest that firms can make use of peer effects, by placing new hires in teams with experienced workers or by providing training only to a particular group of workers instead of all workers. It is not clear, however, whether there is an optimal team composition and which share of workers should be trained for an optimal return to training. It is remarkable to find relatively strong peer effects although the production process is an individual one. If the mechanism behind the peer effects is exchange of information (knowledge spill-overs), then we can expect even higher peer effects in jobs where the production process is organised in a way that workers work more closely together.

Chapters 5 and 6 examine how firms can use incentives to elicit effort from their workers. Chapter 6 shows how bonus payments affect worker performance. Providing well-defined performance thresholds has a positive effect on average worker performance. The results show, however, that these effects are not homogenous across workers. If thresholds are set in a way that workers can reach the threshold without any change in their efforts, the effect can even be negative. The positive effects are particularly strong for workers who can actually reach performance thresholds by increasing their performance. This suggests that absolute performance thresholds have to be set appropriately so that they incentivise all workers.

Even without explicit incentives, however, firms can elicit effort from their workers. The existence of implicit incentives, e.g. through dismissal or promotion decisions, can achieve similar results as performance pay without having the costs of paying bonuses to workers. In Chapter 5, I show that team leaders can increase worker performance by providing ranking information to workers. The results show that here the largest effect is among high-performing agents. This result is in contrast to the results presented in Chapter 6 where I show that explicit incentives rather work at the lower end of the performance distribution. This suggests that implicit and explicit incentives could be complimentary in improving worker effort.

A further implication of this thesis is that performance outcomes that focus on work speed (average handling time) and those on service quality (share of repeat calls, net promoter score) are not necessarily substitutes. At first sight, one may expect that faster work speed would come at the cost of lower quality provided in the worker's task. The results of several chapters of this thesis, however, rather suggest that improvements in average handling time are not associated with lower quality provided and vice versa. This might be due to the fact that for many customers a

7 Conclusion

more efficient shorter call contributes to service quality or that more able agents are simply able to find proper solutions for customers more quickly.

By having four studies on HR practices that focus on learning as well as incentives, one would like to draw conclusions about which of these practices are most important, and how firms should mix them optimally. The way in which these practices can be used, however, depends on the type of firm, the skill requirements of the workers' tasks as well as on the possibility of performance measurement. For generalising the results to other jobs and sectors, it is therefore important to compare the importance of further training and, incentives for worker performance as well as the opportunities for performance measurement.

As a general conclusion, the research of this thesis shows that personnel data allow the researcher to obtain new insights into economic issues that can not be handled adequately by using representative survey data including all sectors of industry. Personnel data can therefore complement these survey data and offer new insights. Apart from direct measurement of workers' task performance, personnel data can open alleys to implement field experiments to achieve identification as well as to analyse peer effects in the work place. This is not only useful for academic research, but also for the firms themselves. Randomly assigning workers to different treatments is the best way to improve firms' human resource management practices.

Bibliography

- AL-UBAYDLI, O., S. ANDERSEN, U. GNEEZY, AND J. A. LIST (2008): “For love or money? Comparing the effects of non-pecuniary and pecuniary incentive schemes in the workplace,” Discussion paper, George Mason University, mimeo.
- ALTONJI, J. G., AND N. WILLIAMS (2005): “Do wages rise with job seniority? A reassessment,” *Industrial and Labor Relations Review*, 58(3), 370–397.
- ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): “Identification and causal effects using instrumental variables,” *Journal of the American Statistical Association*, 91, 444–455.
- ASCH, B. J. (1990): “Do incentives matter? The case of Navy recruiters,” *Industrial and Labor Relations Review*, 43(3), 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(7-8), 435–452.
- BAKER, G., M. GIBBS, AND B. HOLMSTROM (1994): “The internal economics of the firm: Evidence from personnel data,” *Quarterly Journal of Economics*, 109(4), 881–919.
- BANDIERA, O., I. BARANKAY, AND I. RASUL (2005): “Social preferences and the response to incentives: Evidence from personnel data,” *Quarterly Journal of Economics*, 120(3), 917–962.
- (2007): “Incentives for managers and inequality among workers: Evidence from a firm-level experiment,” *Quarterly Journal of Economics*, 122(2), 729–773.

Bibliography

- (2010): “Social incentives in the workplace,” *Review of Economic Studies*, 77(2), 417–459.
- (2012): “Team incentives: Evidence from a firm level experiment,” *Journal of the European Economic Association*, forthcoming.
- BANDIERA, O., V. LARCINESE, AND I. RASUL (2012): “Blissful ignorance? Evidence from a natural experiment on the effect of individual feedback on performance,” mimeo, London School of Economics and Political Science.
- BARANKAY, I. (2011a): “Gender differences in productivity responses to performance rankings: Evidence from a randomized workplace experiment,” mimeo, University of Pennsylvania.
- (2011b): “Rankings and social tournaments: Evidence from a crowdsourcing experiment,” mimeo, University of Pennsylvania.
- BARNETT, A. G., J. C. VAN DER POLS, AND A. J. DOBSON (2005): “Regression to the mean: What it is and how to deal with it,” *International Journal of Epidemiology*, 34(1), 215–20.
- BARRON, J. M., D. A. BLACK, AND M. A. LOEWENSTEIN (1989): “Job matching and on-the-job training,” *Journal of Labor Economics*, 7(1), 1–19.
- BARTEL, A. P. (1995): “Training, wage growth, and job performance: Evidence from a company database,” *Journal of Labor Economics*, 13(3), 401–25.
- BARTEL, A. P. (2000): “Measuring the employer’s return on investments in training: Evidence from the literature,” *Industrial Relations*, 39(3), 502–524.
- BASSANINI, A., A. BOOTH, G. BRUNELLO, M. DE PAOLA, AND E. LEUVEN (2007): “Workplace training in Europe,” in *Education and Training in Europe*, ed. by G. Brunello, P. Garibaldi, and E. Wasmer, chap. 8-13. Oxford University Press, Oxford.
- BATT, R., V. DOELLGAST, AND H. KWON (2005): “Service management and employment systems in U.S. and Indian call centers,” in *Brookings Trade Forum 2005: Offshoring White-Collar Work—The Issues and Implications*, ed. by S. M. Collins, and L. Brainard, pp. 335–372. Brookings Institution Press, Washington, D.C.

- BATT, R., D. HOLMAN, AND U. HOLTGREWE (2009): “The globalization of service work: Comparative institutional perspectives on call centers,” *Industrial and Labor Relations Review*, 62(4), 453 – 488.
- BATTU, H., C. R. BELFIELD, AND P. J. SLOANE (2003): “Human capital spillovers within the workplace: Evidence for Great Britain,” *Oxford Bulletin of Economics and Statistics*, 65(5), 575–594.
- BECKER, G. S. (1962): “Investment in human capital: A theoretical analysis,” *Journal of Political Economy*, 70(5), 9–49.
- BENKARD, C. L. (2000): “Learning and forgetting: The dynamics of aircraft production,” *American Economic Review*, 90(4), 1034–1054.
- BERTRAND, M., E. DUFLO, AND S. MULLAINATHAN (2004): “How much should we trust differences-in-differences estimates?,” *Quarterly Journal of Economics*, 119(1), 249–275.
- BISHOP, J. H. (1989): “On-the-job training of new hires,” CAHRS Working Paper Series 401, Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies, Ithaca, NY.
- BLACK, S. E., AND L. M. LYNCH (2001): “How to compete: The impact of workplace practices and information technology on productivity,” *Review of Economics and Statistics*, 83(3), 434–445.
- BLANES I VIDAL, J., AND M. NOSSOL (2011): “Tournaments without prizes: Evidence from personnel records,” *Management Science*, 57(10), 1721–1736.
- BORGHANS, L., B. GOLSTEYN, AND A. DE GRIP (2007): “Werkend leren,” *Economisch Statistische Berichten (ESB)*, 92, 260–263.
- BREUER, K., P. NIEKEN, AND D. SLIWKA (2012): “Social ties and subjective performance evaluations: An empirical investigation,” *Review of Managerial Science*, forthcoming.
- CARROLL, V. P., H. L. LEE, AND A. G. RAO (1986): “Implications of sales-force productivity heterogeneity and demotivation: A Navy recruiter case study,” *Management Science*, 32(11), 1371–1388.
- CASTILLA, E. J. (2005): “Social networks and employee performance in a call centre,” *American Journal of Sociology*, 110(5), 1243–1283.

Bibliography

- 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(560), 376–399.
- DE GRIP, A., J. SAUERMANN, AND I. SIEBEN (2011): “Tenure-performance profiles and the role of peers: Evidence from personnel data,” IZA Discussion Paper 6164, Institute for the Study of Labor (IZA), Bonn.
- DEARDEN, L., H. REED, AND J. VAN REENEN (2006): “The impact of training on productivity and wages: Evidence from British panel data,” *Oxford Bulletin of Economics and Statistics*, 68(4), 397–421.
- DESTRE, G., L. LÉVY-GARBOUA, AND M. SOLLOGOUB (2008): “Learning from experience or learning from others? Inferring informal training from a human capital earnings function with matched employer-employee data,” *Journal of Socio-Economics*, 37(3), 919–938.
- DOHMEN, T., AND A. FALK (2011): “Performance pay and multidimensional sorting: Productivity, preferences, and gender,” *American Economic Review*, 101(2), 556–90.
- FALK, A., AND A. ICHINO (2006): “Clean evidence on peer effects,” *Journal of Labor Economics*, 24(1), 39–58.
- FAMA, E. (1980): “Agency problems and the theory of the firm,” *The Journal of Political Economy*, 88(2), 288–307.
- FLABBI, L., AND A. ICHINO (2001): “Productivity, seniority and wages: new evidence from personnel data,” *Labour Economics*, 8(3), 359–387.
- GIELEN, A., M. KERKHOFS, AND J. VAN OURS (2010): “How performance related pay affects productivity and employment,” *Journal of Population Economics*, 23(1), 291–301.
- GOUX, D., AND E. MAURIN (2000): “Returns to firm-provided training: Evidence from French worker-firm matched data,” *Labour Economics*, 7(1), 1–19.
- GÜRTLER, O., AND M. KRÄKEL (2012): “Dismissal tournaments,” *Journal of Institutional and Theoretical Economics*, 168(4), 547–562.

- GURYAN, J., K. KROFT, AND M. NOTOWIDIGDO (2009): “Peer effects in the workplace: Evidence from random groupings in professional golf tournaments,” *American Economic Journal: Applied Economics*, 1(4), 34–68.
- HAMILTON, B. H., J. A. NICKERSON, AND H. OWAN (2003): “Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation,” *Journal of Political Economy*, 111(3), 465–497.
- HEYWOOD, J. S., X. WEI, AND G. YE (2011): “Piece rates for professors,” *Economics Letters*, 113(3), 285–287.
- HOLMAN, D., R. BATT, AND U. HOLTGREWE (2007): “The Global Call Center Report: International Perspectives on Management and Employment,” Project report, Global Call Center Project.
- HOLMSTRÖM, B. (1999): “Managerial incentive problems: A dynamic perspective,” *Review of Economic Studies*, 66(1), 169–182.
- HOLMSTRÖM, 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.
- HOLZER, H. J., R. N. BLOCK, M. CHEATHAM, AND J. H. KNOTT (1993): “Are training subsidies for firms effective? The Michigan experience,” *Industrial and Labor Relations Review*, 46(4), 625–636.
- ICHNIOWSKI, C., AND K. SHAW (2012): “Insider econometrics: A roadmap to estimating models of organizational performance,” in *Handbook of Organizational Economics*, ed. by R. Gibbons, and J. Roberts. Princeton University Press.
- ICHNIOWSKI, C., K. SHAW, AND G. PRENNUSHI (1997): “The effects of human resource management practices on productivity: A study of steel finishing lines,” *American Economic Review*, 87(3), 291–313.
- 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.
- KONINGS, J., AND S. VANORMELINGEN (2010): “The impact of training on productivity and wages: Firm level evidence,” IZA Discussion Paper 4731, Institute for the Study of Labor (IZA), Bonn.

Bibliography

- KOSTIUK, P. F., AND D. A. FOLLMANN (1989): "Learning curves, personal characteristics, and job performance," *Journal of Labor Economics*, 7(2), 129–146.
- KRUEGER, A., AND C. ROUSE (1998): "The effect of workplace education on earnings, turnover, and job performance," *Journal of Labor Economics*, 16(1), 61–94.
- LAZEAR, E. P. (1979): "Why Is there mandatory retirement?," *Journal of Political Economy*, 87(6), 1261–1284.
- (1995): *Personnel Economics*, MIT Press Books. The MIT Press.
- (2000): "Performance pay and productivity," *American Economic Review*, 90(5), 1346–1361.
- LAZEAR, E. P., AND S. ROSEN (1981): "Rank-order tournaments as optimum labor contracts," *Journal of Political Economy*, 89(5), 841–64.
- LAZEAR, E. P., K. L. SHAW, AND C. T. STANTON (2012): "The value of bosses," NBER Working Papers 18317, National Bureau of Economic Research.
- LEUVEN, E., AND H. OOSTERBEEK (2008): "An alternative approach to estimate the wage returns to private-sector training," *Journal of Applied Econometrics*, 23(4), 423–434.
- LIU, X., AND R. BATT (2007): "The economic pay-offs to informal training: Evidence from routine service work," *Industrial and Labor Relations Review*, 61(1), 75–89.
- MANSKI, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *Review of Economic Studies*, 60, 531–542.
- MAS, A., AND E. MORETTI (2009): "Peers at work," *American Economic Review*, 99(1), 112–145.
- MEYER, M., AND J. VICKERS (1997): "Performance comparisons and dynamic incentives," *Journal of Political Economy*, 105(3), 547–581.
- MINCER, J. (1958): "Investment in human capital and personal income distribution," *Journal of Political Economy*, 66, 281.
- MORETTI, E. (2004): "Estimating the social returns to higher education: Evidence from longitudinal and repeated cross-section data," *Journal of Econometrics*, 121, 175–212.

- MURALIDHARAN, K., AND V. SUNDARARAMAN (2011): “Teacher performance pay: Experimental evidence from India,” *Journal of Political Economy*, 119(1), 39–77.
- PARENT, D. (2003): “Employer-supported training in Canada and its impact on mobility and wages,” *Empirical Economics*, 28(3), 431–459.
- ROSENBAUM, P. R., AND D. B. RUBIN (1983): “The central role of the propensity score in observational studies for causal effects,” *Biometrika*, 70(1), 41–55.
- SACERDOTE, B. (2001): “Peer effects with random assignment: Results for Dartmouth roommates,” *Quarterly Journal of Economics*, 116(2), 681–704.
- SCHWERDT, G., D. MESSER, L. WOESSMANN, AND S. WOLTER (2012): “Effects of adult education vouchers on the labor market: Evidence from a randomized field experiment,” *Journal of Public Economics*, 96(7-8), 569–583.
- SHAW, K., AND E. P. LAZEAR (2008): “Tenure and output,” *Labour Economics*, 15(4), 705–724.
- SHEARER, B. (2004): “Piece rates, fixed wages and incentives: Evidence from a field experiment,” *Review of Economic Studies*, 71(2), 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(1).
- SIEBEN, I., A. DE GRIP, D. VAN JAARSVELD, AND O. SØRENSEN (2009): “Technology, selection, and training in call centers,” *Industrial and Labor Relations Review*, 62(4), 553–572.
- THORNTON, R. A., AND P. THOMPSON (2001): “Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding,” *American Economic Review*, 91(5), 1350–1368.
- TRAN, A., AND R. ZECKHAUSER (2012): “Rank as an inherent incentive: Evidence from a field experiment,” *Journal of Public Economics*, 96(910), 645–650.
- WILCOXON, F. (1945): “Individual comparisons by ranking methods,” *Biometrics Bulletin*, 1(6), 80–83.
- ZIMMERMAN, D. J. (2003): “Peer effects in academic outcomes: Evidence from a natural experiment,” *Review of Economics and Statistics*, 85(1), 9–23.
- ZWICK, T. (2006): “The impact of training intensity on establishment productivity,” *Industrial Relations*, 45(1), 26–46.

Summary in Dutch

Welke instrumenten kunnen bedrijven inzetten om de prestaties van hun werknemers te verbeteren? Bij de beantwoording van deze vraag staan in dit proefschrift twee soorten instrumenten, die door veel bedrijven worden ingezet, centraal: leren en incentives.

Prestaties van werkenden zijn niet noodzakelijk stabiel over de tijd. Werknemers passen hun inspanning en dus hun prestaties aan aan de prikkels die ze krijgen. Om de prestaties van werknemers, en daarmee de prestaties van het bedrijf te verbeteren, kunnen werkgevers kiezen uit verschillende human resource (HR) instrumenten (bv. Lazear, 1995; Ichniowski et al., 1997). Voorbeelden hiervan zijn instrumenten gerelateerd aan het personeelsverloop, de organisatie van het werk, prestatiebeloning en training.

Leren en incentives zijn beide gericht op prestatieverbetering. Echter, de mechanismen achter beide instrumenten verschillen van elkaar. Leren verhoogt de waarde van het menselijk kapitaal (human capital) van werknemers, met als gevolg een hogere productiviteit. Incentives beïnvloeden daarentegen de inspanning die een werknemer levert, met als beoogd gevolg een hogere productiviteit.

Bassanini *et al.* (2007) rapporteren dat bijna 40% van de werknemers in Nederland tenminste één keer heeft deelgenomen aan een training. Deze trainingen worden meestal gefinancierd door de werkgever. Naast formeel leren via trainingscursussen heeft ook informeel leren (learning-by-doing) een belangrijk aandeel in het verbeteren van de prestaties van werknemers: Borghans *et al.* (2007) tonen aan dat het grootste gedeelte van investeringen in human capital juist plaatsvindt door informeel leren op het werk.

Summary in Dutch

Incentives zijn een essentieel onderdeel van het HR-beleid van bedrijven. Gielen *et al.* (2010) tonen aan dat 39% van de Nederlandse bedrijven gebruik maakt van prestatie-gerelateerde beloning. In callcenters in Nederland is gemiddeld 41% van het loon van adviseurs prestatieafhankelijk (Holman *et al.*, 2007). Naast prestatiebeloning bestaan er ook impliciete incentives, zoals bijvoorbeeld het uitoefenen van druk op werknemers door teamleiders.

Voor het onderzoek van dit proefschrift is gebruik gemaakt van personeelsdata. Personeelsdata bevatten gegevens over medewerkers van een bedrijf (of sector). Een belangrijk kenmerk van de in dit proefschrift gebruikte personeelsdata is de aanwezigheid van informatie over de prestaties van werknemers. Deze informatie komt in representatieve data die verzameld wordt met behulp van vragenlijsten niet voor. De data die in dit proefschrift worden gebruikt zijn verzameld in een callcenter van een internationaal opererend telecommunicatiebedrijf. Het unieke van de verzamelde data is dat er voor alle medewerkers over een lange periode informatie over de prestaties van individuele werknemers beschikbaar is. Daarnaast is er informatie beschikbaar over bijvoorbeeld de duur van het dienstverband van de callcenter medewerkers, de gevolgde trainingen en de indeling in teams. Deze informatie maakt het mogelijk om de hypothesen in dit proefschrift te testen. Hoofdstuk 2 van dit proefschrift beschrijft de data die in het onderzoek wordt gebruikt.

Tenure, peer effects, en prestatie

Hoofdstuk 3 meet de tenure-prestatie profielen van callcenter medewerkers. Bovendien wordt er geanalyseerd in hoeverre deze werkervaringsprofielen de prestatieprofielen van andere werknemers beïnvloeden. De analyse van tenure-prestatie profielen sluit aan bij de literatuur over het effect van de duur van het dienstverband op de hoogte van het loon (Lazear, 1979; Baker *et al.*, 1994; Altonji en Williams, 2005). Een interpretatie van stijgende tenure-loon profielen is dat deze profielen een toename van iemands kennis en vaardigheden (human capital) weerspiegelen. Deze kennis-toename wordt door werkgevers gewaardeerd met hogere lonen (Mincer, 1958). In Hoofdstuk 3 gebruik ik personeelsdata als een alternatieve manier om de relatie tussen de duur van het dienstverband en iemands prestaties te schatten. Zelfselectie van werknemers is bij deze soort data minder een issue. Bovendien bevatten deze data directe objectieve prestatie-maatstaven (Shaw en Lazear, 2008).

De schattingsresultaten laten zien dat er sprake is van steile tenure-prestatie profielen in de eerste maanden dat iemand in het callcenter werkt. Na ongeveer acht maanden wordt dit profiel vlak. Een toename van 10% in de duur van het dienstverband

is gerelateerd aan een toename in prestatie met 5,5% van de standaarddeviatie. Deze toename kan als een investering in informeel leren worden geïnterpreteerd. In het eerste jaar investeert een gemiddeld adviseur 162 uren in dit informele leren. Tijdens dezelfde periode investeert het bedrijf slechts 84 uren in formele trainingscursussen.

Bovendien behandelt Hoofdstuk 3 de vraag of callcenter medewerkers van hun collega's kunnen leren. De resultaten laten zien dat een toename in de ervaring van collega's met één standaarddeviatie samengaat met een toename van het eigen prestatieniveau met 14% van een standaarddeviatie. Dit resultaat toont aan dat het cruciaal is voor bedrijven om nieuwe en ervaren werknemers bij elkaar in een team te plaatsen. Dit faciliteert de kennisspreiding, ook als de teamleden geen gezamenlijke teamprestatie leveren, maar individuele werkzaamheden verrichten.

Het rendement van training programma's

In Hoofdstuk 4 wordt door middel van een veldexperiment het effect van formele training op de prestaties van werknemers geanalyseerd. In dit experiment worden adviseurs willekeurig ingedeeld in een treatment-groep en een controle-groep. De treatment-groep neemt deel aan de training, de controle-groep niet. Doordat adviseurs willekeurig in twee groepen worden ingedeeld kan het causale effect van training op de prestatie van medewerkers worden geïdentificeerd. De identificatie van dit effect is cruciaal om een beter inzicht te krijgen in de rol die training van werkenden heeft voor de ontwikkeling van iemands human capital (e.g., Bassanini *et al.*, 2007). De bestaande literatuur toont aan dat de gebruikte methode en ook de manier waarop productiviteit wordt gemeten een invloed hebben op de grootte van de effecten van een training. De gemeten effecten zijn doorgaans kleiner als er beter wordt gecontroleerd voor zelfselectie (Bartel, 2000; Black en Lynch, 2001; Bassanini *et al.*, 2007; Leuven en Oosterbeek, 2008). Uit het veldexperiment blijkt echter dat callcenter medewerkers in de treatment-groep hun prestaties met 10% verbeteren nadat ze de training hebben gevolgd (direct effect). Maar ook de medewerkers die werkzaam zijn in hetzelfde team als degenen die een training hebben gevolgd profiteren van de training van hun collega's. Ondanks het feit dat zij zelf geen training hebben gevolgd verbeteren de prestaties van deze laatste groep met 0,51%.

Het effect van het uitoefenen van druk op iemands performance

In de meeste organisaties zijn werknemers georganiseerd in teams. Dit vergemakkelijkt het monitoren en evalueren van individuele werknemers. Hoofdstuk 5 analyseert

in hoeverre teamleiders het prestatieniveau van adviseurs kunnen beïnvloeden met behulp van een impliciete incentive, namelijk het uitoefenen van druk.

De bestaande literatuur over de effecten van incentives beschrijft vooral de effecten van expliciete incentives. Literatuur over de effecten van impliciete incentives is schaars. Eén van deze schaarse studies is de studie van Lazear *et al.* (2012) die de rol van managers analyseert. Andere studies zijn de studies van, Blanes i Vidal en Nossol (2011) en Barankay (2011a,b) die aantoont dat het informeren van werknemers over hun positie op de prestatie-ranking een effect heeft op de inspanningen die ze leveren.

Ik meet het effect van het uitoefenen van druk door teamleiders op callcenter medewerkers op zowel individueel niveau als op teamniveau. Ik toon aan dat het effect van de druk om hoger te scoren op de individuele rankings duidelijk sterker is dan het effect van de druk om hoger te scoren op de team rankings. De grootte van dit effect hangt bovendien af van de positie op de ranking. In het bovenste deel van de ranking is dit effect groter dan in het onderste deel van de ranking. Het zijn dus vooral de al beter presterende medewerkers die door de druk die er uitgaat van de rankings nog beter gaan presteren.

Prestatiebeloning en performance

Gebruikmakend van de introductie van prestatiebeloning in het bedrijf waar ik mijn data heb verzameld, analyseer ik in Hoofdstuk 6 het effect van prestatiebeloning op de prestaties van de callcenter medewerkers. Door de prestatiebeloning kunnen callcenter medewerkers maximaal 12% van hun loon als bonus ontvangen. Dit percentage is afhankelijk van het behaalde prestatieniveau.

De resultaten laten zien dat adviseurs inderdaad op de nieuwe incentives reageren. Dit komt overeen met resultaten van studies in andere sectoren (Lazear, 2000; Shearer, 2004; Shi, 2010; Heywood *et al.*, 2011). De effecten doen zich met name voor bij de callcenter medewerkers die het slechtst presteerden voor de introductie van de prestatiebeloning. Voor deze groep zijn de incentives kennelijk het sterkst. De nieuw geïntroduceerde prestatiebeloning heeft daarentegen geen effect op de callcenter medewerkers die al een hoger prestatieniveau hadden.

Biography

Jan Sauermann studied economics at the University of Mannheim with a specialisation in labour economics and econometrics, and graduated in 2004. After working at the Halle Institute for Economic Research (IWH) as research associate, he joined the Research Centre for Education and the Labour Market (ROA) at Maastricht University. Since October 2012, he holds a post-doctoral position at the Swedish Institute for Social Research (SOFI) at Stockholm University. Jan is affiliated to the Institute for the Study of Labour (IZA) in Bonn, Germany, and the Research Centre for Education and the Labour Market (ROA) at Maastricht University.

ROA Dissertation Series

1. Lex Borghans (1993), *Educational Choice and Labour Market Information*, Maastricht, Research Centre for Education and the Labour Market
2. Frank Cörvers (1999), *The Impact of Human Capital on International Competitiveness and Trade Performance of Manufacturing Sectors*, Maastricht, Research Centre for Education and the Labour Market
3. Ben Kriechel (2003), *Heterogeneity Among Displaced Workers*, Maastricht, Research Centre for Education and the Labour Market
4. Arnaud Dupuy (2004), *Assignment and Substitution in the Labour Market*, Maastricht, Research Centre for Education and the Labour Market
5. Wendy Smits (2005), *The Quality of Apprenticeship Training, Conflicting Interests of Firms and Apprentices*, Maastricht, Research Centre for Education and the Labour Market
6. Judith Semeijn (2005), *Academic Competences and Labour Market Entry: Studies Among Dutch Graduates*, Maastricht, Research Centre for Education and the Labour Market
7. Jasper van Loo (2005), *Training, Labor Market Outcomes and Self-Management*, Maastricht, Research Centre for Education and the Labour Market
8. Christoph Meng (2005), *Discipline-Specific or Academic? Acquisition, Role and Value of Higher Education Competencies*, Maastricht, Research Centre for Education and the Labour Market

9. Andreas Ammermüller (2007), *Institutional Effects in the Production of Education: Evidence from European Schooling Systems*, Maastricht, Research Centre for Education and the Labour Market
10. Bart Golsteyn (2007), *The Ability to Invest in Human Capital*, Maastricht, Research Centre for Education and the Labour Market
11. Raymond Montizaan (2010), *Retrenched Pension Rights and Workers' Motivation*, Maastricht, Research Centre for Education and the Labour Market
12. Annemarie Nelen (2012), *Part-Time Employment and Human Capital Development*, Maastricht, Research Centre for Education and the Labour Market
13. Jan Sauermann (2013), *Human Capital, Incentives, and Performance Outcomes*, Maastricht, Research Centre for Education and the Labour Market