The effects of training on own and co-worker productivity: Evidence from a field experiment

Andries De Grip[a,b] Jan Sauermann[a]

February 2011

Abstract

This paper analyses the effects of work-related training on workers' productivity. To identify causal effects from training, we combine a field experiment that randomly assigned workers to treatment and control groups with panel data on individual worker productivity before and after training. We find that participation in the training programme leads to a 9 percent increase in productivity. Moreover, we provide evidence for externalities from treated workers on their untreated peers: An increase of 10 percentage points in the share of treated peers leads to a productivity increase of 0.45 percent. We provide evidence that the estimated effects are causal and not due to selective labour turnover.

JEL-codes: J24, M53, C93
Keywords: Returns to training, field experiment, externalities, peer effects

[a] Research Centre for Education and the Labour Market (ROA), Maastricht University
[b] Institute for the Study of Labor (IZA), Bonn

Corresponding author: Jan Sauermann
Research Centre for Education and the Labour Market (ROA)
Maastricht University
P.O. Box 616, NL–6200 MD Maastricht
Telephone: +31–43–38 83801
j.sauermann@maastrichtuniversity.nl

The authors thank Iwan Barankay, Eric Bonsang, Lex Borghans, Thomas Dohmen, Dan Hamermesh, Ben Kriechel, Olivier Marie, Eric Maurin, Raymond Montizaan, and audiences at Maastricht University, the Education in Adulthood and the Labour Market-Workshop (IAB Nuremberg), the IWAEE-Workshop (Catanzaro), EEA 2010 (Glasgow), RWI Essen, and the Advances with Field Experiments workshop (Wharton School) for helpful comments. This paper has been screened to ensure no confidential information is revealed.

1 Introduction

Estimating the returns of work-related training on productivity is a challenge for two reasons. 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 (cf. 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 rare.

This study exploits a field experiment to measure the causal effects of investments in training on worker productivity. The field experiment was carried out in the call centre of a multinational telephone company operating 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 productivity 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 which affect both training participation and productivity. We find that agents are on average 9 percent more productive after participating in the training programme than non-treated agents. We show that the effects are causal and not the result of employee selection. Workers belonging to the treatment group were trained over successive weeks. This time-varying treatment, furthermore, allows us to identify possible externalities of training on untreated peers. We find that an increase of 10 percentage points in the share of treated agents leads to an increase in performance of 0.45 percent.

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 two main strands.[1] The first 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; Konings and Vanormelingen 2010) or direct measures of productivity within one sector of industry (Holzer et al. 1993). 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 firms' production processes (Ichniowski and Shaw 2009). Moreover, it is difficult, if not impossible, to find direct measures of productivity that are comparable across industries.

The second strand of research on the effects of training on productivity has focused on just a single firm or establishment, or comparable establishments within a sector of industry (Bartel 1995; Ichniowski et al. 1997; Krueger and Rouse 1998; Liu and Batt 2007). Overall, previous studies have found a range of estimated treatment effects of training participation on productivity. At the same time, there is consensus that appropriate correction for selection into training matters for estimation.

The externalities of training have hardly been discussed in the literature on the impact of training on productivity (cf. Dearden et al. 2006). Recent literature, however, has shown evidence of social effects between workers that may operate either through human capital spillover, peer pressure, or prosocial behaviour (Falk and Ichino 2006; Kato and Shu 2008; Guryan et al. 2009; Mas and Moretti 2009; Bandiera et al. 2010; De Grip et al. 2011). Training externalities have important implications for investments in further training. If training has positive externalities, firms can benefit from targeting a subgroup of workers for training instead of training all workers.

[1] Studies on the wage returns to training have been extensively surveyed in Bartel (2000) and Dearden et al. (2006).

This study aims to contribute to the human capital literature on the effects of work-related training on worker productivity. First, to overcome selectivity in training participation, we exploit the random assignment to training by means of a field experiment. Controlling appropriately for the endogeneity of training participation is crucial, since ordinary least squares estimates are potentially biased due to selection on unobservable characteristics. Previous studies have used different methodologies to take selectivity into account, such as panel approaches to control for time-invariant unobservable factors, and instrumental variables. Although the use of field experiments has sharply increased over the last decade, no studies in the human capital literature have yet exploited exogenous variations in training participation to estimate the effects of training.[2] By randomly assigning agents to treatment and control groups, we can estimate the causal effect of training on individual productivity. Additionally, we use detailed administrative information on agents working in the call centre to show that our estimates are not affected by worker selection into training.

Second, we explicitly model the possibility of training externalities that may operate either through human capital spillover or peer effects from social pressure. The literature on human capital externalities originates from growth theory and has been, among others, applied to human capital spillover from education (Moretti 2004). If the returns to training are estimated at the individual (worker) level without taking knowledge spillovers into account, returns will be underestimated. This implication also holds for the second type of externalities, peer effects due to social pressure (Falk and Ichino 2006). The identification of externalities, however, is empirically difficult. Previous studies used either quasi-random variations in group composition (Mas and Moretti 2009) or random assignments of subjects to groups (Guryan et al. 2009). While many field experiments have exploited changes in management behaviour induced by the firm's management (e.g.

[2]In the literature on the wage effects of training, are two studies are noteworthy: Leuven and Oosterbeek (2008) used information on the reasons for non-participation to artificially create treatment and control groups *ex-post*, whereas the study of Schwerdt et al. (2011) used randomly assigned vouchers to estimate the local average treatment effect (LATE) on wages. Neither study found any significant treatment effects.

Shaw and Lazear 2008), we randomly split teams into training groups, allowing us to identify the causal estimates of within-team externalities.

Third, we use panel data on individual workers' productivity to estimate the effects of the training programme. To measure productivity, we use the key performance indicator used by the call centre for evaluating its call agents, that is the average time needed to handle inbound customer calls. This productivity measure has also been used in other studies ion the call centre sector (Liu and Batt 2007; Breuer et al. 2010). We measure each worker's productivity each week before and after the training, allowing us to analyse short-run productivity dynamics after training. Using individual productivity as an outcome allows us to capture the total effect of the training on productivity, while wage information captures only the share an agent receives.[3]

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) estimated that call centres employed about 4 million employees, roughly 3 percent 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 technologies in in-house centres requires high investments in work-related training (Sieben et al. 2009).

The remainder of this paper is structured as follows: Section 2 provides an overview of the firm analysed in this paper and its workers and describes the experiment in detail. Section 3 presents our regression model and the estimation results. Additional evidence and tests on the robustness of our results are discussed in Section 4. Finally, Section 5 summarises the paper and presents concluding remarks.

[3]A second advantage over 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.

2 Context of the experiment

2.1 Organisation of the department and worker tasks

The field experiment analysed in this paper was implemented in an in-house call centre of a multinational 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 for private customers with fixed cell phone contracts.

The main task of call agents in this department is to answer phone calls from customers calling the firm. Customers contact the 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 on promotional campaigns, communication, and information technology skills, as well as courses on handling more complex calls.[4]

In the first period of our sample, 179 individuals were working in the department. Column (1) of Table 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 19.3, and only 7.6 percent of agents are working 30 hours or more each week. A total of 29.1 percent of all agents are men, the average age of the agents is 32.6 years, and the agents have on average 2.6 years of experience working for the firm.

Agents are organised in 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

[4]Some more specific training courses allow the management to assign more calls of one type to an agent who has been trained in this type of calls. However, the training analysed in this paper does not prepare for handling specific types of calls. The estimation of the training programme's treatment effect is thus not affected by a changing composition of calls.

work on the same floor of the building. Work places are organised as '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 productivity scorecards containing detailed information on key performance indicators. While information on the average length of calls is easy to quantify, it is more difficult to monitor agent call quality. Team leaders could obtain additional information on agent quality by listening in on their calls, from customer satisfaction surveys, or from the repeat call rate, which is used as a proxy for call quality.

Agent pay is based on a single collective agreement. Agents are paid a base wage and receive an annual increase depending on an annual performance rating by the team leader. Based on this grading, agents receive an annual wage raise of 0 to 8 percent. There are no performance-related bonuses.

2.2 Training purpose, contents, and organisation

The training programme analysed in this study was intended for all agents of the department. Its aim was to increase the efficiency of agents answering customer calls. Management had decided to organise the training to decrease the average time needed to handle calls because the call centre was performing below its targets.

The training was organised as a week-long programme of 38 hours. 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. 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 techniques designed to decrease the average time needed for handling customer calls. This included for instance, the way in which call agents gathered information from the customers in order 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.

2.3 The field experiment

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 study uses exogenous variations in training participation to identify the causal effect of training on individual productivity.

As shown in Figure 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 period; 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 are thus never directly treated throughout our observation period.

Out of the 179 individuals working in the department at the beginning of our observation period (week 50/2008), 86 were selected for the experiment. The aim of this selection was to minimise the number of dropouts due to high turnover among recently hired agents. Management thus primarily selected by agent tenure. Column (4) of Table 1 shows the differences between assigned and non-assigned agents. Assigned agents are on average six years older, have a longer tenure by more than three years, and are more productive. 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 the control groups is exogenous conditional on being assigned to the training programme. An advantage of the pre-selection by management is that agents are relatively homogeneous. Agents with longer tenure deal with all types of calls, which is less likely for those who just started their job.

Conditional on being selected for participation in the training, 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. The differences in observable characteristics between the assigned treatment and control groups are relatively small, with none significantly different from zero (Column (3) of Table 2). Due to selection, three agents from the treatment group and nine agents from the control group do not enter the estimation sample (Columns (4) to (6) of Table 2). The differences between the resulting treatment group and the control group is not significantly different from zero.[5]

Because of the size restriction of training groups, eight of the ten teams had to be split into separate training groups. The training groups of one team were assigned to either the treatment group or the control group. In between the training weeks of a team's training groups, a group from a different team was trained. During this week, the team consisted of exogenously chosen agents who were treated and exogenously chosen agents who were not yet treated. We exploit the fact that the agents of one team were randomly assigned to training groups to identify within-team externalities. Table 3 shows

[5]In 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.

the means of observable variables for agents who are in a team's first training group versus the means for agents who are in a team's second training group. Except for the measure of call quality, there are no differences in means significantly different from zero.

All agents were informed about the training at the moment of their initial assignment. Throughout the experiment, information about the randomisation of the order of teams was only given 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 for this information, since the agents are used to teams being split and separately trained because of the training centre's limited capacity.

Throughout our main analysis, we only use the compliers for our estimations. The data is an unbalanced panel of 74 agents during 32 weeks. In total, we have 1,859 agent-week observations on productivity. On average, agents work in 28 weeks out of the 32 weeks.

2.4 Measuring productivity

This study measures agent productivity by means of the call centre's key performance indicator: the average handling time, which is defined as the average time an agent needs to handle a customer call (cf. Liu and Batt 2007, Breuer et al. 2010). We use the average handling time of agent i in week t as a measure of productivity. Information on the average handling time of individual agents is available for all working weeks. 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.

The department's aim is to improve performance by decreasing average handling time. Throughout this study, we 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 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 0.381, which relates to 4.4 minutes 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. Compared to individual heterogeneity in productivity, however, variation over time is less important: While period fixed effects explain only about 11 percent of the overall variation in outcomes, worker fixed effects alone explain 52 percent of the overall variation in individual productivity.

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 first must 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, however, does not violate the assumption that our measure of productivity 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 productivity 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.

3 Empirical analysis

3.1 Causal effect of training on individual productivity

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

$$y_i = y_i(d_i) = d_i \cdot y_i(1) + (1 - d_i) \cdot y_i(0) \tag{1}$$

where $y_i(1)$ and $y_i(0)$ denote productivity 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 average treatment effect τ (ATE, Rosenbaum and Rubin 1983) is thus identified by $ATE = \tau = E[y_i(1) - y_i(0)]$. The ATE τ can be estimated by performing a linear regression of individual i's productivity 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:

$$y_{it} = \alpha_i + \gamma_k + \tau d_{it} + \beta_1 t_t + \beta_2 X_{it} + u_{it} \tag{2}$$

where α_i are individual fixed effects to account for individual heterogeneity remaining despite the experimental design. X_{it} are covariates that are assumed to be independent from the the treatment status d_{it}, such as working hours in week t.[6] The variable t_t is a linear time trend that controls for trends in aggregate productivity affecting all agents, and u_{it} is an idiosyncratic error term. Throughout this analysis, standard errors are clustered at the agent level.[7]

[6]We 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 has shown that experience effects matter only for new agents (De Grip et al. 2011).

[7]In 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.

3.2 Baseline results

We provide first evidence of the treatment effect in Figure 2, which shows the average productivity of agents in the treatment group. The treatment week is denoted as week 0, with positive (negative) values of the x axis showing the tth calendar week after (before) the training. Productivity appears to be, on average, higher in the weeks after the training than before the training. When the productivity of untreated agents is not controlled for, treated agents perform significantly better after the training. Given the random assignment of agents to the treatment and control group, 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 shows that agents in the treatment group are significantly more productive after the training, while there were no significant pre-treatment differences between the two groups.

Table 5 shows the results when estimating Equation (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 being zero, the share of treated agents increases in time with the growing number of groups that have been trained. Column (1) shows that agent productivity after participation in the training is 10.9 percent higher than before the training, controlling for untreated agents' productivity. When, in addition, individual heterogeneity is controlled for by including worker fixed effects, the effect increases slightly to 12.5 percent (Column 2). Figure 2, however, shows that aggregate trends seem to matter. When controlling for a linear time trend, the effect decreases by about 4.5 percent to an estimated effect of 8.1 percent (Column 3 of Table 5).[8]

The preceding analysis made 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 data at hand, we can exploit dynamic patterns

[8] To allow for more flexibility in time trends common to all agents, we estimated Equation (2) including also linear and squared time trends. The estimated treatment effect $\hat{\tau}$ then becomes slightly lower. The difference between the two estimates is not significantly different from zero.

of the treatment effect. Figure 3 shows the estimated shape of the treatment effect in the weeks after the training. Figure 3 is based on an estimation of Equation (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 productivity in the first week after the training is not significantly different from zero. In the weeks thereafter, agent productivity is significantly higher than before the training. Soon after, however, the estimated treatment effect 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. To better approximate how the training affects productivity, we set the treatment dummy in the first post-treatment period to zero and re-estimate Equation (2). As expected, the average post-treatment performance is higher than the results shown in Table 5 (9.2 percent compared to 8.1 percent). Column (2) of Table 6, furthermore, shows that when interacting the treatment variable with a (individual) post-treatment trend, the shift is higher but decreasing each week.[9] The shift is 13.5 percent, with each additional week; however, the performance relative to untrained individuals decreases by 0.7 percent. The overall treatment effect in the second week is thus 12.1 percent.

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. 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 productivity will increase when their coworkers 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

[9]The post-treatment trend is calculated as being zero in an untreated state, and increasing by one starting from the training week.

the control group are not actually fully untreated but indirectly affected by the training as well.

3.3 The role of social interaction

The tasks agents carry out are individual tasks. There are thus no externalities inherent in the production technology, although, agents are exposed to other agents by spatial proximity. There are reasons to believe that social interactions arise and may thus impact the treatment effect. Agents belonging to the same team sit next to each other. Furthermore, they have the possibility to communicate with their co-workers during team meetings or breaks. Therefore, the likelihood of externalities is higher within teams than across teams. We do not preclude that externalities work between teams. However, given the physical distance on the work floor, the externalities between teams should at least be smaller than externalities within teams.[10]

The preceding subsection estimates the ATE from participation in the training programme as is done in the training literature. An unbiased estimate $\hat{\tau}$, however, requires the stable unit treatment value assumption or 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 that the control group's productivity ($y(0)$) 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:

$$y_i = y_i(d_i) = d_i \cdot y_i(1) + (1 - d_i) \cdot (1 + d_i^* \tau_s) \cdot y_i(0) \tag{3}$$

where d_i, $y_i(0)$, and $y_i(1)$ are defined as in Equation (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

[10] In a survey held among the agents, 80 percent of them stated that they exchange work-related information either during official team meetings or at their workplace.

exposed to treated workers from the same team. If an agent is untreated ($d_i = 0$) and working in a team of untreated agents ($d_i^* = 0$), the potential outcome from Equation (3) simplifies to that from Equation (1). The same holds for agents who have been treated ($d_i = 1$ 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 (3) implies that, unless $\tau_s = 0$, the estimated treatment effect ($\hat{\tau}$) is underestimated in Equation (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 exploit the fact that team agents were randomly assigned to 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. This random assignment enables us to identify within-team externalities. Because positive externalities from treated agents should lead to increased performance even when an agent is untreated, we can test whether $\hat{\tau}$ is lower for agents who were exposed to treated agents before their own training.

Table 7 shows the results of estimating Equation (2) separately for the two samples. Column (1) shows the estimated treatment effect for agents who comprise the first group of their team to be treated, compared to 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 (9.4 percent) than when estimating Equation (2) (8.1 percent; see Column (3) of Table 5). This suggests that the causal effect from training was 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 (2) of Table 7 shows that their treatment effect is 5.7 percent. The difference between the two estimates, however, is not significantly different at the 5 percent significance level.

The results of a second, more direct test of externalities are shown in the third column of Table 7. Here, we test whether the share of treated peers matters in 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 that agent worked. The share of treated peers is calculated for peers from the same team only. In order not to confound the results with the actual treatment, the sample is restricted to agents who were not yet treated. The results show that a 10 percent increase in the share of treated peers leads to a performance increase of .45 percent. Although only statistically significant at the 10 percent level of significance, the results shown in Table 7 strongly suggest the existence of peer effects from treated agents on untreated peers in their team.

3.4 Returns to training

Standard theory predicts that firms invest in human capital if the expected returns from training investments exceed costs. Through direct access to company information, it is straightforward to calculate the costs of the training programme, both in total and per call agent. Quantifying the returns to the training, however, is more complicated. We define the costs of training one agent as:

$$C = \frac{(n_g \cdot w_a + w_l + w_c) * 38}{n_g} \tag{4}$$

where n_g is the size of the training group, and w_a, w_l, and w_c are the hourly wages for the agents, team leader, and coach, respectively. Because our measure of productivity 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 work load and the agent's wage:

$$B = (\hat{\tau} \cdot aht_{it} \cdot n_{c,it}) \cdot w_a \tag{5}$$

where $n_{c,it}$ is the number of calls an agent completes in week t, and B is thus the decrease in the time a trained agent needs to handle a certain workload compared to an untrained agent, multiplied by his wage.

Given the costs of a full week of training and the decreasing treatment effect over time, we use the average performance before the start of the training to calculate the number of weeks a treated agent needs to work at the new level of (improved) performance to recoup the costs of the training. This calculation shows that for an estimated treatment effect of 9 percent, agents need to work 8.8 weeks at their improved performance level to recoup their training costs.[11]

4 Additional evidence

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. We can rule out the argument that this is due to selectivity, since over 90 percent of the agents who participated in the training also worked the following week. A likely explanation, however, is fatigue. Most of the agents work part-time, with an average of about 20 hours per week. The training programme, however, is one fulltime 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 the same, irrespective of the call agent's usual working hours (see Column (1) of Table 8). It is different, however, for long training programmes. For such programmes, 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.

[11] Because half of the training consisted of actual customer calls, we calculated the costs of the training using 50 percent of the agents' wage costs. Considering the full wage costs of agents roughly doubles the recoupment period.

4.2 Personnel turnover and training effects

As in other call centres, the call centre analysed in this study experiences a high agent turnover. In total, we assigned 86 agents to the treatment and control groups, of which 12 left the department before the training. Table 2 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. While we find that agents initially assigned to the treatment and control groups did not differ with respect to observables, the 12 departing call agents may have been selective in terms of unobservable characteristics.

OLS estimation of Equation (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 measure of productivity, average handling time. This type of selectivity occurs when agents with these characteristics are more likely to drop out of the treatment group from out of the control group.

Of the 74 agents of our estimation sample, 55 eventually followed the training. Of the 34 agents assigned to the treatment group, 26 (76 percent) were trained. For the control group, out of the 40 agents assigned, 29 (73 percent) complied with the assignment. The main reason for the higher sample attrition in the control group is that those agents followed the training several weeks later, which increases the probability of dropping out. Calculating the attrition rate for treatment group agents who were staying at least until the last week of our observation period ($N = 36$), 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.[12]

[12] There 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.

To analyse the effect of selection on the estimated treatment effect $\hat{\tau}$, we estimate Equation (2) for two different samples. Our sample used in the analysis so far uses the sample of agents who were randomly assigned to the treatment and control groups and complied with their assignment, that is they did not switch from the treatment to the control group or vice versa. Their treatment effect, shown in Column (1) of Table 9, is 8.1 percent. When we exclude agents who did not stay until the end of our observation period, the number of agents in total decreases by 19, to 55 agents. The estimated treatment effect thereby decreases to 6.8 percent, which is lower than for the bigger sample but not significantly different from the estimated treatment effect for the main sample. These results suggest that agents who selected out of the control group were, on average, of lower ability than agents who selected out of the treatment group.

4.3 Effects on call quality

Throughout this study, we used a transformation of the average length of calls as a measure of actual productivity y_{it}. However, one can argue that y_{it} is a function of the average handling time *and* the call quality q_{it}: $y_{it} = f(aht_{it}, q_{it})$. If y_{it} is a non-monotonic function in aht_{it}, improvements in average handling time will not necessarily translate into higher productivity. Since call quality is more difficult to monitor than work speed, agents may aim to improve their handling times by providing lower call quality.

To test this hypothesis, 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 to which an agent talked to who called the call centre again within seven days. This measure should, on average, give information on whether the agent was able to solve the customer's problem or answered the client's question satisfactorily. On average, 20.2 percent of the customers called back within seven days after the first phone call (see Column (1) of Table 1). Because low values of repeat calls (rc_{it}) are assumed to indicate higher performance, we generate a call quality measure of productivity $y^q_{it} = \frac{1}{rc_{it}} * 10$, with an average of 4.95.

The first column of Table 10 shows that the argument that short calls are associated with lower call quality is not supported by the data. The correlation between the main measure of productivity y_{it} and the call quality measure y_{it}^{q} is only -2.25. When estimating the effect of the treatment on y_{it}^{q}, we do not find that call quality is affected by the training suggesting that agents do not substitute quality with quantity.

In addition to the information on the share of repeat calls, we use data gathered in a customer-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 we know the week in which the call was made and the corresponding agent, 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 (2) through (4) of Table 10 indicate show the estimation results. The results indicate that the training had a positive effect on the agent's knowledge and understanding capabilities. However, according to Column (4), the training had no effect on the agent's ability to provide a solution to the problem.

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 matter for 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 are more productive 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 higher effects from participation in training for more experienced agents. Column

(1) of Table 11 shows the treatment effect for agents with a tenure above the median tenure in our sample. Compared to Column (2) which shows regression results for agents with a tenure below the median, the estimated treatment effect is slightly higher for experienced agents (by 2.2 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 11 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 more frequently have contracts with shorter working hours.

5 Conclusions

This study 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 assigned agents to training groups, thereby generating exogenous variation in training participation. Regression results show that agent productivity of agents was about 9 percent higher after having followed the training. Furthermore, we show that agents who did not participate in the training also improved their performance through externalities. These indirect effects may arise either due to knowledge spillover between trained and untrained workers or social pressure. We find that increasing the share of treated peers by 10 percentage points leads to an increase in performance of 0.45 percent. 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 As the tasks carried out by call agents follow an individual production technology, these externalities are probably even higher in other sectors of industry in which team work is more important than in the call centres.

Our findings also have 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 optimal not to train all workers, or to have different skill training programmes for different parts of the work force.

References

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.

Bandiera, O., I. Barankay, and I. Rasul (2010): "Social Incentives in the Workplace," *Review of Economic Studies*, 77(2), 417–459.

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.

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.

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.

Breuer, K., P. Nieken, and D. Sliwka (2010): "Social Ties and Subjective Performance Evaluations: An Empirical Investigation," IZA Discussion Paper 4913, Institute for the Study of Labor (IZA), Bonn.

De Grip, A., J. Sauermann, and I. Sieben (2011): "Tenure, spill-over effects, and output," Maastricht University, mimeo.

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.

Falk, A., and A. Ichino (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 39–58.

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.

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 (2009): "Insider Econometrics: Empirical Studies of How Management Matters," NBER Working Papers 15618, National Bureau of Economic Research, Cambridge, MA.

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.

KATO, T., AND P. SHU (2008): "Performance Spillovers and Social Network in the Workplace: Evidence from Rural and Urban Weavers in a Chinese Textile Firm," IZA Discussion Paper 3340, Institute for the Study of Labor (IZA), Bonn.

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.

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.

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.

Moretti, E. (2004): "Estimating the Social Returns to Higher Education: Evidence from Longitudinal and Repeated Cross-Section Data," *Journal of Econometrics*, 121, 175–212.

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.

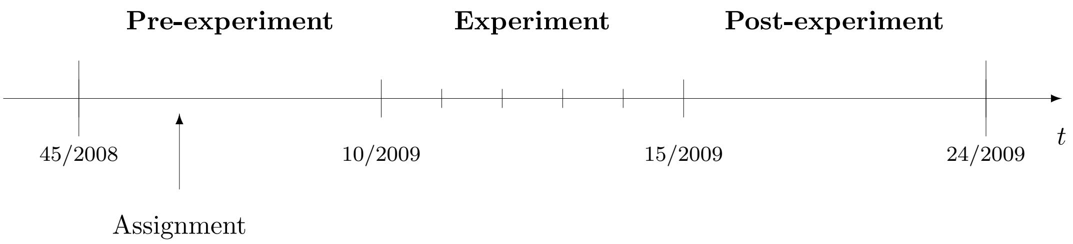
Schwerdt, G., D. Messer, L. Woessmann, and S. Wolter (2011): "Effects of Adult Education Vouchers on the Labor Market: Evidence from a Randomized Field Experiment," IZA Discussion Papers 5431, Institute for the Study of Labor (IZA), Bonn.

Shaw, K., and E. P. Lazear (2008): "Tenure and Output," *Labour Economics*, 15(4), 705–724.

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.

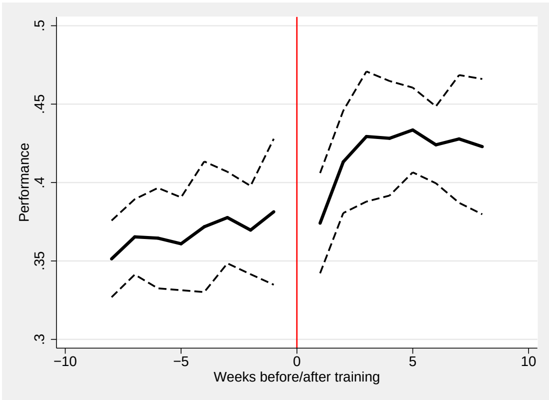
Wilcoxon, F. (1945): "Individual Comparisons by Ranking Methods," *Biometrics*, 1, 80–83.

Figure 1: Overview field experiment



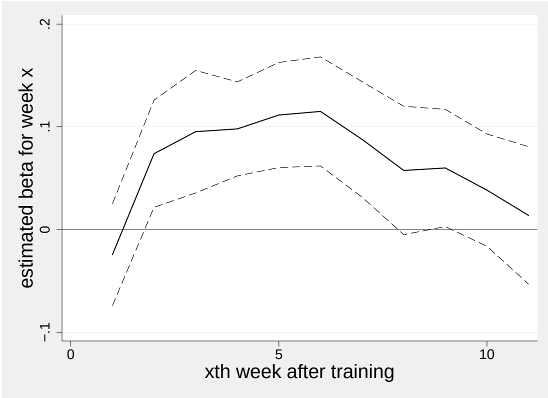
A Figures

Figure 2: Average productivity of the treatment group before and after the training



Note: This figure shows the average productivity (see Section 2.4 for a definition) of agents in the treatment group in the weeks before and after the training week (solid line). The dashed lines show the 95 percent confidence interval estimated from a regression of productivity on week dummies . The training week is denoted as week 0.

Figure 3: Development of treatment effect on after average handling time over time



Note: Development of the treatment effect on after average handling time over time. This figure shows the estimated treatment effect on productivity (see Section 2.4 for a definition) during each week after the training, controlling for a linear time trend. Week 0 denotes the training week. The dashed lines show the 95 percent confidence interval estimated from a regression of productivity on week dummies.

B Tables

Table 1: Selection into the experiment

Agents:	(1) All	(2) Assigned	(3) Non-assigned	(4) Difference(3)-(2)
Gender	.2905	.2790	.3011	-.0220
(share of male agents)	(.4553)	(.4512)	(.4612)	
Age	32.60	35.582	29.431	6.143***
	(11.33)	(11.06)	(10.79)	
Tenure	2.627	4.290	1.089	3.200***
(in years)	(3.610)	(4.019)	(2.307)	
Productivity	.3272	.3752	.2828	.0924***
	(.1060)	(.0807)	(.1076)	
Call quality	.4731	.4558	.4884	-.0326*
	(.1070)	(.0862)	(.1211)	
Average working hours	19.330	18.65	19.96	-1.306
	(10.261)	(9.14)	(11.21)	
Number of agents	179	86	93	

Difference significant at * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; two-sided t test. Standard deviations are shown in parentheses. The number of dropouts is defined as the number of agents who left during our observation period. Descriptive statistics are calculated before agents were assigned to the training (week 50/2008).

Table 2: Descriptive statistics treatment and control group

	Initial assignment (N=86)			Estimation sample (N=74)		
	(1)	(2)	(3)	(4)	(5)	(6)
Agents:	TG	CG	Diff TG-CG	TG	CG	Diff TG-CG
Gender	.3513	.2244	.1269	.371	.275	.096
(share of male agents)	(.4840)	(.4216)		(.490)	(.452)	
Age	34.25	36.58	-2.332	34.47	36.15	-1.680
	(10.02)	(11.79)		(10.22)	(11.64)	
Tenure	4.206	4.353	-.147	4.172	3.754	.418
(in years)	(3.793)	(4.219)		(3.825)	(3.986)	
Productivity	.3687	.3801	-.0114	.3654	.3746	-.0091
	(.0847)	(.0781)		(.0836)	(.0686)	
Call quality	.4519	.4587	-.0069	.4548	.4633	.0086
	(.0583)	(.0161)		(.0569)	(.1121)	
Average working hours	17.51	19.51	-1.996	17.94	20.43	-2.48
	(9.386)	(8.954)		(9.42)	(9.18)	
Number of agents	37	49		34	40	

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

Table 3: Descriptive statistics of teams' first and second training groups (treatment group only)

	(1)	(2)	(3)
Agents:	First group	Second group	Difference (3)-(2)
Gender	.3846	.3636	.0210
(share male agents)	(.5064)	(.4924)	
Age	35.66	33.76	1.90
	(10.76)	(10.07)	
Tenure	3.688	4.458	-.770
(in years)	(3.979)	(3.796)	
Productivity	.3541	.3720	-.0180
	(.0859)	(.0836)	
Call quality	.4797	.4382	.0415**
	(.0543)	(.0538)	
Average working hours	21.84	15.64	6.209
	(8.54)	(9.33)	
Number of agents	13	21	

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).

Table 4: Average performance post-treatment period (61 agents)

	(1) Treatment group	(2) Control group	(3) Difference TG-CG
Average performance (predicted)	.4148	.4027	.0122***
Standard error	(.0012)	(.0012)	(.0017)
Number of agents	28	35	
Number of observations	296	409	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Predicted performance is generated from a regression of performance on working hours of agent i in week t plus a full set of week fixed effects. Standard errors were clustered at the agent level.

Table 5: Direct treatment effect (74 agents)

	(1)	(2)	(3)
Treatment dummy	.1092***	.1248***	.0805***
	(.0293)	(.0153)	(.0205)
Working hours	.0002	-.0022*	-.0024**
	(.0018)	(.0011)	(.0012)
Time trend			.0027***
			(.0009)
Constant	-1.0105***	-.9643***	-1.1520***
	(.0468)	(.0230)	(.0604)
Individual fixed effects	No	Yes	Yes
Observations	1859	1859	1859
Number of agents	74	74	74
R^2	.0320	.5802	.5880

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: log(productivity). All standard errors are clustered at the agent level.

Table 6: Shape of the treatment effect (74 agents)

	(1)	(2)
Treatment dummy	.0916***	.1353***
	(.0216)	(.0186)
Interaction post-training trend × treatment		-.0068***
		(.0019)
Working hours	-.0022*	-.0022*
	(.0012)	(.0012)
Time trend	.0026***	.0028***
	(.0009)	(.0009)
Constant	-1.1444***	-1.1577***
	(.0590)	(.0595)
Individual fixed effects	Yes	Yes
Observations	1859	1859
Number of agents	74	74
R^2	.5902	.5916

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: log(productivity). All standard errors are clustered at the agent level. In both regressions, the treatment dummy is set to zero for the first week *after* the training.

Table 7: Estimation of externalities

	(1)	(2)	(3)
Treatment dummy	.0942***	.0572**	
	(.0213)	(.0282)	
Share of peers treated			.0454*
			(.0270)
Working hours	-.0026*	-.0021	-.0020
	(.0015)	(.0014)	(.0012)
Time trend	.0032***	.0030***	.0027***
	(.0009)	(.0009)	(.0010)
Constant	-1.1757***	-1.1717***	-1.1544***
	(.0703)	(.0647)	(.0665)
Individual fixed effects	Yes	Yes	Yes
Observations	1352	1510	1554
Number of agents	53	61	74
Number of agents (treatment group)	13	21	34
Number of agents (control group)	40	40	40
R^2	.5962	.5549	.5681

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: log(productivity). 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 (3) is restricted to agents in an untreated state only (treatment dummy==0).

Table 8: Pre-/post-differences in performance for other training programmes

Length of courses	(1) short	(2) long
Agents with short working hours	-.045	-.0373
	(.020)	(.079)
Number of observations	744	42
Agents with long working hours	-.039	-.048***
	(.034)	(.012)
Number of observations	64	527

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The numbers show the differences 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.

Table 9: Estimated treatment effect and estimation samples

	(1)	(2)
Treatment dummy	.0805***	.0680***
	(.0205)	(.0211)
Working hours	-.0024**	-.0043***
	(.0012)	(.0010)
Time trend	.0027***	.0035***
	(.0009)	(.0008)
Constant	-1.1520***	-1.1590***
	(.0604)	(.0646)
Individual fixed effects	yes	yes
Observations	1859	1532
Number of agents	74	55
Number of agents (treatment group)	34	26
Number of agents (control group)	40	29
R^2	.5880	.5872

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable: log(productivity). All standard errors are clustered at the agent level. The samples are defined as follows: (1) comprises all agents randomly assigned and complying with the treatment status; (2) is the sample of (1) minus agents dropping out due to selection.

Table 10: The effect of training participation on quality

Outcome	log(Call quality)	Grade 1	Grade 2	Grade 3
	(1)	(2)	(3)	(4)
Treatment dummy	-.0022	1.3484***	2.1704***	.2189
	(.0397)	(.2710)	(.2913)	(.6820)
Working hours	.0021*	.0061	.0149	-.0490
	(.0011)	(.0314)	(.0225)	(.0499)
Time trend	-.0001	-.0121	-.0008	.0192
	(.0021)	(.0174)	(.0214)	(.0240)
Constant	-.8451***	7.4651***	5.5488**	6.3908**
	(.1688)	(2.0806)	(1.9177)	(2.3242)
Individual fixed effects	yes	yes	yes	yes
Observations	1648	112	115	107
Number of agents	73	14	15	15
R^2	.5020	.0050	.0092	.0122

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All standard errors are clustered at the agent level. Call quality is measured as 1/(share of repeat calls*10). 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).

Table 11: Direct treatment effect with interaction terms

Outcome	Below-median tenure	Above-median tenure	
	(1)	(2)	(3)
Treatment dummy	.0761***	.0894***	.1543***
	(.0248)	(.0317)	(.0368)
Working hours			-.0019
			(.0012)
Working hours × treatment			-.0036***
			(.0013)
Constant	-1.1640***	-1.2205***	-1.1586***
	(.0796)	(.0868)	(.0592)
Individual fixed effects	yes	yes	yes
Observations	908	951	1859
Number of agents	36	38	74
R^2	.0969	.0766	.0953

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column (1) and (2) show the baseline regression for agents with below-median (1) and above-median (2) tenure, respectively. Dependent variable: log(productivity). All standard errors are clustered at the agent level.