The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

CESifo Working Papers
www.cesifo.org/wp

Reciprocity in Organisations
Evidence from the WERS

Florian Englmaier
Thomas Kolaska
Stephen Leider

CESIFO WORKING PAPER NO. 5168
CATEGORY 13: BEHAVIOURAL ECONOMICS
JANUARY 2015

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5168*

Reciprocity in Organisations
Evidence from the WERS

Abstract

Recent laboratory evidence suggests that social preferences may affect contractual outcomes under moral hazard. In accordance with previous research, this paper uses written personality tests for job candidates as a proxy for whether firms care about personality traits of employees, in particular whether these employees are inclined towards reciprocity. Using the British Workplace Employment Relations Survey 2004 (WERS) we find that behavior of employers and employees is consistent with the presence of gift-exchange motives: firms that screen applicants for personality are more likely to pay generous wages and to provide (non-pecuniary) benefits like employer pension, on-the-job training, or job security. Firms likewise benefit from reciprocal employees as they can implement more team-working and are generally more successful. Other modern human resource practises like competency tests or incentive pay only poorly predict these patterns. Moreover, there is no association between dismissals and personality tests, indicating that personality tests do not merely improve the fit between applicant and employer. Hence, we conclude that motivation based on gift-exchange motives remains as the most plausible explanation for our results.

JEL-Code: D220, M520.

Keywords: reciprocity, organisational structure, employee compensation.

Florian Englmaier
Department of Economics
University of Munich
Munich / Germany
florian.englmaier@econ.lmu.de

Thomas Kolaska
Department of Economics
University of Munich
Munich / Germany
thomas.kolaska@econ.lmu.de

Stephen Leider
University of Michigan
Ross School of Business
USA - 48109-1234 Ann Arbor MI
leider@umich.edu

July 16, 2014
We are grateful to Sascha Becker for helpful discussions. The second author thanks the University of Warwick's Economics Department, where part of this research was completed, for its hospitality. Financial support from the Deutsche Forschungsgemeinschaft through GRK 801 (Kolaska) is gratefully acknowledged.

1 Introduction

Understanding the behaviour of employees in labour relations is crucial for managers and firm owners who aim to align potentially diverging interests of management and workforce. In the last decades contract theorists developed a consistent framework in which monetary incentives induce agents to exert effort, serving as a guideline for real-world firms.[1] While modern human resource departments (to some extent) rely on theoretical considerations, the majority of real-world labour contracts are characterised by fixed payments and – if at all – only a minor part of employees' income is attributed to incentive pay.[2] The classic static moral hazard theory would, absent explicit incentives, predict lower levels of effort exertion than the real-world examples show.

Incorporating concepts from behavioural economics may provide additional explanations for these real-world observations. In an early contribution, Akerlof (1982) demonstrates that wages may exceed the market-clearing wage when employers attempt to influence working norms via gift-exchange.[3] More recently, Englmaier and Leider (2012) introduce the concept of reciprocity – i.e. gift-exchange motivation – into the classical principal-agent framework concluding that firms with reciprocal employees have more leeway to cost-efficiently induce effort: shifting away from direct monetary incentives and inducing employees to behave reciprocally towards their employers allows firms to save high costs from risk premia they would have to pay when using strong incentives. Relying on reciprocity, however, requires firms to screen for employees with reciprocal traits.

In this paper we use Englmaier and Leider (2012) as a theoretical guideline and search for evidence for the use of reciprocity based motivation in organisations. Using the 5th wave of the "Workplace Employment Relations Survey" (WERS 2004) a large scale survey of Britain-based firms, we find evidence for firm behaviour consistent with gift-exchange motivations. We interpret the use of compulsory personality tests for job candidates as an indicator whether firms explicitly screen applicants for personality traits that may be correlated with job candidates' inclination towards reciprocity. In line with gift-exchange motives, these firms are more likely to provide their employees high wages and other non-pecuniary benefits like employer pension schemes and extended paid annual leave.

[1] See Prendergast (1999) for a survey.

[2] Lemieux et al. (2009) estimate that approximately 37% of male labour market participants in the US (using the PSID, Panel Study of Income Dynamics 1976 - 1998) receive variable payments with a median magnitude of 3.5%. Englmaier and Leider (2012) discuss further studies corroborating this argument.

[3] Also in a labour market context, Becker et al. (2013) provide field evidence for heterogeneous long-term responses to gift-exchange motivation.

Furthermore, employees in these establishments enjoy more on-the-job training (c.f. Leuven et al. (2005)) and have a higher chance that their employer provides guaranteed job security.

Screening applicants' personality and providing benefits for those who get hired may pay off for the firm if employees reciprocate with higher effort. Even though we are not able to measure effort directly, we find that employers using personality tests report higher levels of firm performance and are more likely to organise work in teams. The latter is particularly interesting as team work is one of the key modern HR practices and is considered to increase productivity by reaping synergies (Barton et al. (2003)). To allow for effective team work, team members need to subordinate their own desires to the common good, i.e. ought to be reciprocally cooperative.

In contrast, two additional measures for the presence of modern human resource practises within a firm – competency tests for job candidates and variable payments for employees – fare much worse in predicting benefits for employer or employees. This implies that only screening for personality as opposed to the use of competency tests or other human resource practises explains patterns consistent with gift-exchange motives. The lack of an association between personality tests and dismissals within an establishment furthermore indicates that personality tests do not merely increase the "fit" between employer and employee, which otherwise might have caused similar relationships between screening and benefits. Moreover, including competency tests for job candidates and variable payments for employees in the analysis also serves as control such that we do not merely pick up the general sophistication of a firm's human resource policies in our regressions.

Closest to this work is Huang and Cappelli (2010). Based on a national survey of US employers, they argue that employers who state that they particularly value applicants with high "work ethic" are less prone to monitor their employees, organise more work in teams, and have lower turnover rates. Furthermore, employees in firms looking for motivated employees receive higher wages and these firms are more productive. Comparing their results with ours we can, by and large, confirm their findings, with the exception that we do not find any relationship between personality tests and turnover and monitoring respectively.[4]

While we regard our study as complementary to Huang and Cappelli (2010), we are distinct in at least two main dimensions: First, the richness of the WERS allows us to

[4]Institutional differences like unionisation rates (in the US 12 - 13 percent and in the UK about 29 percent in 2002/2003 (Lesch (2004) and Visser (2006)) between the United States and Britain however do not allow for direct comparisons of the results.

include the entire range of occupational groups within an establishment, from managers to unskilled labour, into large parts of our analysis. In contrast Huang and Cappelli (2010) use data on frontline workers only.[5] Both studies, however, have in common that the cross-sectional structure of the data does not allow to pin down a unique explanation of the observed pattern. Despite evidence in favour of reciprocity as the underlying principle we cannot establish causality.

Second, Huang and Cappelli (2010) use a survey question in which managers have to rate how important candidates' "work ethic" is for them when assessing applicants. In contrast, we use "hard" information on whether written personality tests are used in the hiring process. These tests are based on observable practises, implying that other datasets may contain this measure as well which ensures that the analysis is transferable to other data sources containing information on test use. Using personality tests as a measure for reciprocity within the labour force, the difference in the interpretation of the results between Huang and Cappelli (2010) and our work becomes clear: we favour – based on a fully fledged agency model in Englmaier and Leider (2012) – the mechanism of reciprocity as conditional intrinsic motivation as a plausible explanation of our observations, whereas "work ethic" refers to unconditional motivation of the employee. However, both explanations are – with the data at hand – observationally similar. Hence, the additional evidence in favour of reciprocity in this paper is still only of suggestive nature.

Personality tests are only one potential dimension of how firms screen job candidates. Other popular methods are interviews, reference letters and – widely used – competency tests.[6] Whereas the latter aims to uncover cognitive ability, personality tests – the "Big Five" framework is a prominent example – measure a whole range of characteristics of a potential employee. In particular we interpret the use of these personality tests as a proxy for firms that are more likely to have (highly) reciprocal workers. In many cases this may be due to screening for other desirable traits that are correlated with reciprocity, though firms may be also directly screening for reciprocally cooperative types.

Our empirical approach is consistent with findings that document that personality traits usually identified with personality tests within the "Big Five" framework are (closely) correlated with measures of *reciprocity* as commonly defined in laboratory experiments. In a real world setting, Autor and Scarborough (2008) document the hiring procedures

[5] Within the WERS dataset we are unable to restrict our estimations to frontline workers because the majority of the dependent variables in question is available only on firm level or for the largest occupational group.

[6] See Rynes and Cable (2003) for an extensive review on the various methods employed in modern hiring procedures.

of a large retail firm – which according to the authors is representative for the industry – that uses personality tests to screen workers upon hiring. The firm gave hiring preference to applicants with positive z-scores for "agreeableness", "conscientiousness", and "extroversion", "Big Five" traits that are predictive for the presence of reciprocity. Ben-Ner et al. (2004) link behaviour in a dictator game with switching roles to previously elicited personality traits and find that "Big Five" indicators "agreeableness" and "openness" are associated with higher amounts a dictator sends in response to the amount she previously received. Opposed to that, cognitive ability seems not to influence the propensity to reciprocate. In an earlier contribution, Ashton et al. (1998) concludes on basis of hypothetical questions that high "agreeableness" and high "emotional stability" are associated with high reciprocal altruism. Englmaier and Leider (2010) provide also evidence from a laboratory experiment that "Big Five" indicators correlate with subjects' behavior in a gift exchange task. Consistent with the patterns in our data, Autor and Scarborough (2008) provide evidence that firms widely make use of screening for personality and Wilk and Cappelli (2003) show that employers differ substantially in the extent to which they make use of applicant screening.

The importance of social preferences for individual decisions has been documented in various studies, for extensive surveys see Fehr and Schmidt (2003) and for field evidence DellaVigna (2009). Fehr and Gächter (2000) in their survey explicitly concentrate on the prevalence of reciprocity. In several theoretical contributions, social preferences have been associated to optimal contract designs, suggesting that not only productivity and ability but also social traits can influence the generosity of contract offers.[7] In an empirical study using survey data, Dohmen et al. (2009) provide evidence from real-world labour markets for the importance of reciprocity on wages and effort provision. Englmaier et al. (2014) in a real-effort laboratory experiment elicit both productivity and social preferences from agents and find that principals increase wages for both traits by adapting contract offers accordingly. In an earlier contribution, Cabrales et al. (2010) predict outcomes in a gift-exchange experiment on basis of elicited behavioural preferences.

Another strand of the personnel literature explores synergies between different human resource practises. Using firm data from steel finishing lines, Ichniowski and Shaw (1997) find that the use of modern human resource practises – like incentive payments, work being organised in teams, flexible job assignment, job security, and training for employees – is positively associated with productivity of these firms. In a recent experimental study Bartling et al. (2012) find complementarities between high discretion, high wages and rent sharing, job characteristics which are commonly associated with "good jobs". The

[7] See for references Itoh (2004), Dur and Glazer (2007) or Englmaier and Wambach (2010)

authors demonstrate that these jobs emerge endogenously (as they are profitable) if employers have the opportunity to screen job candidates. Importantly, they show that it is screening for social preferences and not for competency which is necessary for "good jobs" to emerge.

We contribute to the literature by combining evidence from both strands of the literature. We show that screening applicants for personality is associated with a bundle of benefits for employees and employers. In contrast, firms' use of competency tests fails to predict these outcomes. Moreover, personality tests are unrelated to dismissals within firms. Hence, explanations solely targeting on correlations between successful firms and application of modern human resource practises are too narrow and we feel confident that our results point to a more nuanced, behavioural, explanation of the observed patterns: The systematic use of reciprocity based motivation by firms.

The remainder of the paper is structured as follows: In Section 2, we provide an extensive description of the WERS 2004 with a special focus on personality tests. Section 3 contains details on the Hypotheses, the estimation strategy, and results. Furthermore a substantial part of the section is dedicated to robustness checks. Section 4 concludes with a discussion.

2 Data

2.1 The WERS 2004 Dataset

The empirical analysis relies on the 2004 "Workplace Employee Relations Survey" (WERS 2004), the fifth in a government-funded series of surveys carried out at British workplaces.[8] The WERS 2004 covers information on employment relations of British workplaces and is provided by employees and employers. The following analysis entirely relies on the information about establishments provided by employers.[9] This dataset is weighted using standard weights to account for the sampling design and is stratified according to the suggested procedure which is a combination of number of employees and industry code.[10]

[8] For further information on the WERS see: http://www.wers2004.info.

[9] The WERS consists of different datasets with varying respondents. Besides the survey answered by employers, the WERS also comprises datasets on employees and employee representatives with questions targeted to figure out their individual view on the establishment and their working conditions. The latter two datasets are not employed for our analysis because they do not contain structured information that would aid our strategy. Moreover, we lack information on the selection process of a workplace's employees for this survey part implying potential endogeneity problems.

[10] For reference, see http://www.wers2004.info/FAQ.php#stata, section 5.6 "How do I apply weights and correctly estimate variances in Stata?", April 23, 2014.

The WERS 2004, consisting of 2,295 establishments surveyed, is a representative sample of the British economy.[11] The number of employees per establishment varies widely between a minimum of 5 jobs per workplace up to 10,006 with an average of 414 jobs per workplace. Note, however, that the mean is inflated by few extremely large companies – the median firm size is 69 jobs and even the 99th percentile only contains a maximum of 4,936 jobs per workplace. The firms cover almost all branches of the economy with a slight concentration on health, whole trade and manufacturing.[12] About one fourth of the establishments are attributed to the public sector. More than half of the establishments are unionised (58 percent).

21 percent of the establishments are part of the productive sector, three quarters are one of a number of different workplaces in the United Kingdom belonging to the same organisation, 23 percent are a single independent establishment and 2 percent are the sole UK establishment of a foreign company. Overall 78 percent of the firms are either entirely or predominantly UK-owned, whereas the controlling head office of the company is foreign-based in only 12 percent of the cases. Market shares are widely dispersed with approximately 39 percent (15 percent) of the firms indicating a market share of less than five percent (more than 50 percent). Roughly in line with this, about 75 percent of the firms report that the perceived degree of competition in their market is either high or very high whereas 11 percent state it to be low or very low.[13]

Within each firm the WERS 2004 distinguishes between 9 different occupational groups.[14] Panel (d) of Figure 1 provides absolute frequencies for all nine occupational groups pooling all 2,295 establishments. Not surprisingly, almost all firms state to have a management department and about 80 percent of the surveyed firms have employees in secretarial or administrative positions. As several variables of interest, including modern human resource practises, are provided on occupational group level, our subsequent analysis relies on both, firm level and occupational group level.

Table 1 provides summary statistics for variables of interest, including the following

[11] "WERS 2004 (...) provide(s) a nationally representative account of the state of employment relations and working life inside British workplaces." Source: http://www.wers2004.info/wers2004/wers2004.php, October 23rd 2012.

[12] Workplaces are classified according to the SIC 2003 (Standard Industrial Classification) by the UK National Statistics. Sectors not covered by the WERS 2004 include: Agriculture, hunting and forestry, fishing, mining and quarrying, private households with employed persons, and extra-territorial bodies.

[13] The fractions of the legal state, market share and the degree of competition are calculated dropping any unclear answers.

[14] These occupational groups are: (1) Managers and senior officials, (2) professional occupations, (3) associate professionals and technical occupations, (4) administrative and secretarial occupations, (5) skilled trade occupations, (6) caring, leisure and other personal services, (7) sales and customer service occupations, (8) process, plant, and machines operatives, and drivers, and (9) routine and unskilled occupations.

Table 1: Summary Statistics

	Obs.	Avg.	SD	Pctl. 25	Pctl. 50	Pctl. 75	Min.	Max.
				Firm Level				
Monitoring	2278	2.52	1.08	2	2	3	1	7
Dismissal	2160	0.02	0.05	0	0	0.01	0	1.3
Firm Performance	2160	0.54	0.5	0	1	1	0	1
Firm Benefit	2295	0.8	0.4	1	1	1	0	1
Firm Benefit 2	2295	0.89	0.31	1	1	1	0	1
Job Security	2295	0.17	0.37	0	0	0	0	1
Top Wage	2135	0.18	0.24	0	0.08	0.27	0	1
Low Wage	2135	0.03	0.13	0	0	0	0	1
Personality Tests	2292	0.34	0.47	0	0	1	0	1
Competency Tests	2291	0.61	0.49	0	1	1	0	1
Incentive Pay	2295	0.57	0.49	0	1	1	0	1
				Largest Occupational Group				
Possibly Non-Pecuniary Benefits								
Any Benefit	2286	0.89	0.31	1	1	1	0	1
No. Benefis	2286	2.6	1.36	2	3	3	0	5
Pension Scheme	2286	0.77	0.42	1	1	1	0	1
Company Car	2286	0.18	0.38	0	0	0	0	1
Private Health	2286	0.20	0.40	0	0	0	0	1
Extended Paid Leave	2286	0.75	0.43	1	1	1	0	1
Sick Pay	2286	0.70	0.46	0	1	1	0	1
On-the-Job Training	1950	4.05	1.09	3	4	5	1	6
General Training	2288	0.58	0.49	0	1	1	0	1
Team-working	2279	5.08	2.25	3	6	7	1	7

Notes: Statistics for each variable are calculated omitting "refusal", "don't know" and "not applicable", indicating unclear answers. "Job Security", "Personality Tests", "Competency Tests" and "Incentive Payments" are collapsed on firm level to guarantee comparability. The lower panel refers to information on the largest occupational group only. "Monitoring", "On-the-Job training", and "Team-Working" are ordinal variables with lower values corresponding to lower levels of monitoring, training, and team-working respectively. "Dismissal", "Top Wage", and "Bottom Wage" are continuous fraction of dismissed employees, and employees with high and low earnings. "No. Benefits" counts the number of granted benefits, the remaining variables are binary.

statistics: the number of observations, averages and standard deviations, the 25th, 50th and 75th percentile as well as minimum and maximum values. The first set of variables is reported on firm level.

"Monitoring" is an ordinal variable asking for the proportion of non-managerial employees who have job duties which involve supervising other employees. Value one indicates that no employee has monitoring tasks – on average firms indicate that between one

and 19 percent of the workers have monitoring tasks. The continuous variable (relative) "Dismissal" measures the percentage of the workforce which has been dismissed within the previous year. The data suggest that dismissals occur very rarely.

"Firm Performance" is an indicator which combines the following self-reported performance measures: "Financial Performance", "Labour Productivity" and "Product Quality". The indicator is one, if firms in at least one of the three dimensions report to have better performance than the median answer of all firms for each dimension.[15] This classification based on self-reporting, splits the data into two almost equal parts of rather successful and unsuccessful establishments.

We construct the variable "Firm Benefit" as a comprehensive measure of success of establishments. It either relates to self-reported outcomes or to the ability of the firm to employ highly desirable work practices: The measure takes the value one if the respective firm either reports higher than median firm performance, uses more team-working than the median firm, or relies less on monitoring (compared to the median).

We also use an alternative indicator for overall firm benefits, "Firm Benefit 2" which includes dismissals and reports high benefits if additionally the firm has dismissals lower or equal to median dismissals. The purpose of this procedure is to fully address all firm benefits – monitoring, team-working, dismissals, and productivity – which were suggested by Huang and Cappelli (2010) in one compound measure. However this procedure comes at a price: By doing so, we lose much of the variation as most firms do not have any dismissals within the previous year, c.f. Table (1), resulting in almost 90 percent of firms being classified as firms which reap some suggested benefits.

The data contain rich information on various aspects of the workers compensation and benefit package. "Job Security" is reported for each occupational group. In this table, however, we collapse the measure on firm level. Hence the dummy variable on "Job Security" is one if at least in one occupational group within an establishment employees enjoy job security or non-compulsory redundancies.[16] Finally, "Top Wage" ("Low Wage") is an indicator variable which provides information about the relative size of top-wage (low-wage) earners compared to all employees within a firm. The WERS defines the highest wage category (we label this category "Top Wage") as wages equal or more than

[15] For these three measures employers were asked to rate the performance of their firm compared to the relevant industry, resulting in heavily over-rated own performance: for the example of "Labour Productivity", 49 percent of employers state to be better or a lot better than the average and 94 percent state to be at least about average for the industry. Overrating is similarly severe for variables "Financial Performance" and "Product Quality". To account for this overrating we classify establishments as successful if their own rating is better than the median rating of all firms.

[16] Non-compulsory redundancies cover voluntary redundancies and early retirement, see https://www.gov.uk/staff-redundant/noncompulsory-redundancy, November 20, 2012.

15 pounds per hour. The dataset provides three more wage categories: 4.5 or below ("Low Wage" category), 4.51 – 5, 5.01 – 15 or 15 and above pounds per hour.

For the second set of variables the dataset provides measures for the largest occupational group in terms of employees. Possibly non-pecuniary benefits for the employee comprise different measures of benefits for a worker. "Any Benefit" is a binary variable indicating whether employees of the largest occupational group receive any of five benefits suggested in WERS 2004, with almost 90 percent of firms providing at least one benefit.[17] "Number of Benefits" is an ordinal measure how many (between zero and five) different benefits of the suggested five benefits employees receive. We furthermore provide summary statistics for all in the survey suggested benefits, namely "Pension Scheme", "Company Car", "Private Health Insurance", "Extended Paid Leave" and "Sick Pay".

The variable "On-the-Job Training" is measured ordinally with value one indicating that employees of the largest occupational group did not experience any training within the previous year and six implying ten days and more. The WERS 2004 furthermore distinguishes between providing training on computing skills, team-working, communication skills, leadership skills, operation of new equipment, customer service, health and safety, problem-solving methods, equal opportunities, reliability and working to deadlines and quality control procedures. We classify team-working skills, communication skills, and leadership skills under the label "General Training" as these cover matters which are not narrowly job-specific but may be considered as more general and hence can be beneficial for an employee's entire working life across different employers.

Finally "Team-working" is an ordinal variable asking for the proportion of employees in the largest occupational group being designated to teams. No team-working at all (value one) is rather rare, and an average of almost three indicates that 60 – 80 percent of largest occupational group employees work in teams.

2.2 Modern Human Resource Practises

The WERS 2004 provides detailed information about human resource practises within establishments including the prevalence of personality tests, competency tests and variable payments by occupational group. In order to be able to control for the fact that some workplaces might have an in general more sophisticated HR department we classify establishments that use these three practices as employing modern human resource practises.

[17] The survey asks for the following non-pay terms and conditions: (1) Employer pension scheme, (2) company car or car allowance, (3) private health insurance, (4) more than four weeks paid annual leave, (5) sick pay in excess of statutory requirements, and (6) none of these.

More than one third of all establishments use personality tests when screening job candidates whereas more than 60 percent of firms make use of competency tests in at least one occupational group (see Table 1). Both personality tests and competency tests are less prevalent in sectors with lower skill intensive tasks (i.e. construction, wholesale and retail, and hotels and restaurants) while we find high rates of competency tests in financial services, public administration and education.[18] Similarly, personality tests are prominent in financial services, public administration and manufacturing. A correlation coefficient of $\rho = 0.02$ ($\rho = 0.05$) between self-rated perceived degree of competition and personality (competency) tests provides little evidence for enhancing effects of market pressure on the introduction of modern human resource practises.

Analysing the prevalence of both screening tests within the firm it is no surprise that screening devices are most common for hiring managers. Excluding managers, in about 24 percent (56 percent) of establishments personality (competency) tests are required at least in one occupational group when recruiting new employees.

The prevalence of personality and competency tests by occupational group is summarised in detail in panel (a) and (b) in Figure 1. Comparing both panels it again becomes clear that employers use competency tests more often when hiring applicants: for each occupational group the relative frequency of competency tests exceeds that one of personality tests. More interestingly, the distributions of both tests differ to a large extent. Whereas firms make use of competency tests to a similar extent across occupational groups (with exceptions of personal services and unskilled labour with clearly lower rates) the prevalence of personality tests starkly declines with decreasing skill intensity. The exception is the group of sales employees who are very likely to be screened for personality upon hiring.

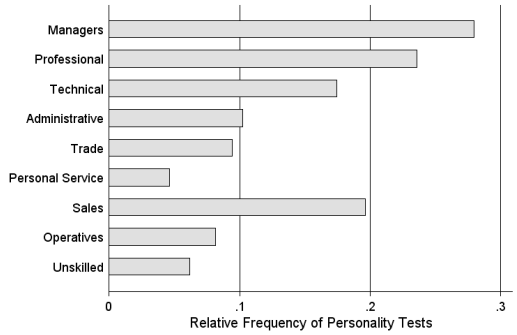
A comparison of both distributions provides some tentative evidence that personality tests and competency tests are measuring different characteristics of the job candidate and are applied to different job requirements. This assessment is further supported by a correlation coefficient of only $\rho = 0.24$ between personality tests and competency tests implying no strict path dependency in firms' choice of which screening devices to apply.[19] Of all firms, 38 percent only screen for competency and 5.5 percent exclusively screen for personality upon hiring, whereas 39 percent apply both devices.[20]

[18] One notable exception of lower skill intensive tasks and very high rates of both tests is the sector "electricity, gas and water" as classified by the UK National Statistics. However this may not be representative due to small sample size of only 45 observations for this sector.

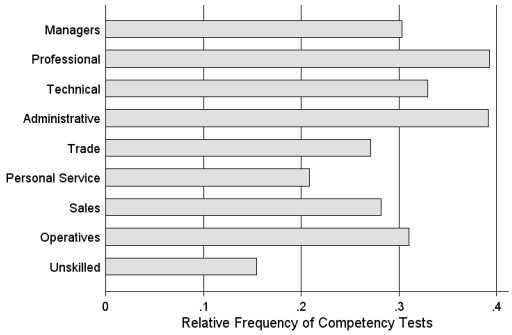
[19] This measure correlates personality tests and competency tests both for all occupational groups excluding managers.

[20] Including managers shifts these fractions a bit: 38 percent of firms use exclusively competency tests, 7 percent personality tests and 39 percent both tests.

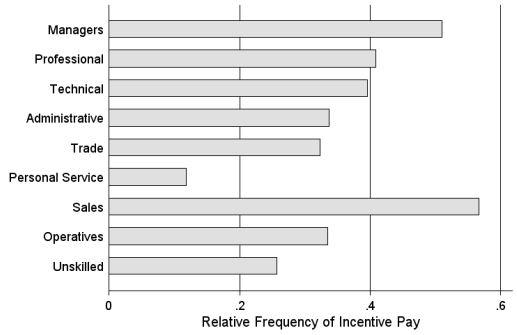
Figure 1: Relative Frequency of Human Resource Practises by Occupational Group
This figure provides an overview over the prevalence of personality tests, competency tests (upper row) and incentive pay for nine different occupational groups (bottom left). The picture on the bottom right depicts the prevalence of each of the nine occupational groups in absolute terms.



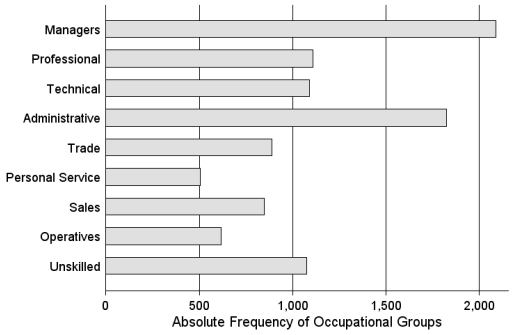
(a) Personality Tests



(b) Competency Tests



(c) Incentive Pay



(d) Occupational Group

The third measure for advanced human resource practises are incentive components in employees' compensation schemes. Paying some sort of variable payment – either performance pay or profit pay – is common in 57 percent of establishments (and in half of the firms in at least one occupational group if abstracting from the group of managers).[21] As can be seen from panel (c) in Figure 1 the distribution of incentive pay across occupational groups declines for less skill intensive tasks with the exception of sales, where incentive pay is common.

[21]The dataset only indicates whether a firm provides variable pay for a certain occupational group but does not give estimates of its magnitude compared to the fixed wage.

3 Reciprocity in Organisations

In this Section we use the presence of personality tests in a firms' hiring procedure as a proxy for this firm having a weakly more reciprocally inclined workforce. Even if a firm does not use personality tests to directly screen for reciprocal workers, it might end up with a more reciprocal workforce as a by-product (see the discussion in the Introduction). This allows us to test various hypotheses regarding reciprocity in organisations within one data set. Some of these hypotheses have independently been advanced in Leuven et al. (2005), Huang and Cappelli (2010), and Englmaier and Leider (2012).

3.1 Hypotheses

The model in Englmaier and Leider (2012) serves as a loose theoretical background for developing the following hypotheses. In this model employers can employ incentives based on gift-exchange if two conditions are fulfilled: First, in a labour market with heterogeneous agents, the employer has to screen for reciprocal job candidates, willing to repay a generous contract offer with increased effort. Second, the willingness of these reciprocally inclined employees to reciprocate needs to be "activated" by the employer via initial "kind behaviour". More technically, the employer has to offer a contract that exceeds the agent's outside option. This can be achieved by offering a higher than market wage[22] or, as the model is based on utility arguments, by providing other, possibly non-pecuniary, benefits, like an employer pension scheme or paid annual leave.

According to Leuven et al. (2005), firms with a more reciprocal workforce are more likely to provide training to their employees. Besides regarding training as additional benefit for workers, reciprocal behaviour of the agent might be a necessary condition for the provision of on-the-job training. As benefits from training are inherently sequential, the employer has to trust her employee that the employee does not enjoy the training and then leaves for a better offer. Put differently, training could be regarded as an increase in the worker's outside option. Furthermore if the employer is convinced of the worker's reciprocal behaviour, she may be willing to provide relatively more general training, which is advantageous not only for a specific job but for the worker's entire employment biography.

In a similar vein, employers may provide job security to their labour force, signalling confidence in workers' loyalty towards the firm. If agents, however, lack reciprocal at-

[22] A different explanation of high wages is provided by Huang and Cappelli (2010): As workers with high work ethic help the firm to save costs, employers attempt to hire as many of these types as possible, which drives up their wages (rent sharing).

tachment to the establishment, job security schemes enable employees to exploit this device via shirking while being protected against immediate consequences.

Hypothesis 1 (Generosity to Workers) *Firms which screen for personality pay higher wages, are more likely to provide their workforce additional (potentially non-pecuniary) benefits, should have a higher likelihood to provide their workers higher amounts of on-the-job training, in particular more general training, and should be more inclined to provide job security to their employees.*

On the other hand, making use of motivational devices which are based on reciprocal behaviour is costly for firms in the first place, as most gifts like higher wages or pension systems involve direct costs. Job security, for instance, inhibits employers to adjust the size of the labour force to fluctuations in demand in the short run. Hence, a rational employer using reciprocal motivation should expect to enjoy some benefits which at least offset these investments. Though the data does not allow us to pin down the cost-efficiency of a firm's behaviour, we proceed in our analysis by providing some insightful correlations.

Screening job candidates for their personality may be associated with employers' inclination to organise tasks in teams. If firms benefit from team-working under the condition of non-shirking and it is harder to measure effort of each team member compared to individual production (as the employer may only observe team output) then the implementation of team structures should be more likely in organisations with more reciprocal employees. Hence we regard the option for firms to use team-working if necessary as a benefit which can (more easily) be achieved with reciprocal workers. This leads us to the hypothesis that organisations with compulsory personality tests and team-working of employees should be complements.

The strongest link between reciprocity and benefits for the firm are correlations of firm performance and screening job candidates for personality. Such relationships could imply that firms relying on reciprocity as a means of motivating workers on average are more successful in the market.

Huang and Cappelli (2010) document correlations between "work ethic" monitoring and turnover respectively. First, they argue that screening for "work ethic" and monitoring should be substitutes as employees with high "work ethic" exert effort voluntarily. Second, turnover decreases because the fit between job candidate and the firm should be better – a classical matching argument. For completeness, we include these two claims into the set of our testable hypotheses.

Hypothesis 2 (Value to Firms) *Firms which screen for personality should have more leeway to organise tasks in teams and should perform better in the market.*

3.2 Methods

To study the correlations between reciprocity and different outcome variables we use personality tests upon hiring as a measure for reciprocity. The general specification of our estimations is the following reduced form model:

$$y_{id} = P_{id}\beta_P + \mathbf{I'_{id}}\beta_{\mathbf{I}} + \mathbf{X'_i}\beta_{\mathbf{X}} + \epsilon_{id}$$

where y_{id} is the outcome of the dependent variable in occupational group d of firm i. The subscripts of P_{id}, an indicator for the use of personality tests, are defined accordingly. $\mathbf{I_{id}}$ are indicators, which are (for each establishment) available on occupational group level. $\mathbf{X_i}$ are (firm-wide) firm fixed controls and ϵ_{id} is an error term, which is assumed to be i.i.d. across firms but may be arbitrarily correlated within firms (between occupational groups). These potential within-firm correlations are accounted by clustering on firm-level.

Estimations differ in two main dimensions: First, we distinguish whether the dependent variable is reported at the firm-level or separately for each occupational group. The latter allows for matching between human resource practises and dependent variables on occupational group level. Second, different y_{id} are scaled differently, suggesting to adapt estimation strategies accordingly.

Job security – i.e. non-compulsory redundancies as defined in footnote 16 – is the only outcome variable which is provided for each occupational group; hence we estimate the effect of personality tests on job security pooling all available occupational groups.[23] This implies that we are able to match the provision of job security for each occupational group with the employed set of modern human resource practises.

The next set of dependent variables only contains information on the largest occupational group within an establishment. This set consists of all non-pecuniary benefits for the employee, including "Pension Scheme" and "Extended Leave", the "No. Benefits" as well as its prevalence ("Any Benefits"). Furthermore it comprises "On-the-

[23] For this estimation we use all occupational groups per firm and create an indicator, whether the firm provides job security for employees in the respective occupational group. As differently sized firms may have more of less occupational groups (and hence giving firms with more groups a higher number of observations) and different sized firms may at the same time be differently likely to provide their employees job security, we include the number of occupational groups per establishment into the set of controls. By this procedure we aim to reduce the likelihood that this effect may confound the results.

Job Training", "General Training" and "Team-working". We adapt the model accordingly and replace the dependent variable with the outcome for the largest group $y_{i\widetilde{d}_i}$, $\widetilde{d}_i = max(\#\text{ of employees}(d_i))\ \forall i$ and $d \in \{\text{professionals, ..., unskilled occupations}\}$. We proceed analogously for personality tests $P_{i\widetilde{d}_i}$ and $\mathbf{I_{i\widetilde{d}_i}}$. Firm fixed controls which are summarised in $\mathbf{X_i}$ are unaffected.

For the remainder of the dependent variables, i.e. "Dismissal", "Monitoring", "Low Wage", "High Wage", "Firm Performance", and "Firm Benefit" the dataset only provides information at the firm level and lacks individualised occupational group specific data. Hence we construct aggregate measures from the occupational specific measure of personality tests, defining indicator P_i being one if in at least one occupational group (excluding managers and senior officials) job candidates are screened via personality tests.[24] Analogously to personality tests, we collapse $\mathbf{I_{id}}$ to the firm level and obtain $\mathbf{I_i}$.

Secondly, as outcomes are reported on different scales for different variables we adapt estimators accordingly. "Low Wage", "High Wage" and "Dismissal" are continuous variables, suggesting OLS estimation. Both employee benefits "Pension Scheme" and "Extended Leave" as well as the indicator whether the firm pays any benefits ("Any Benefit") are binary outcomes, implying probit regressions. The same applies to the variables "Job Security", "General Training", "Firm Performance", and "Firm Benefit". Finally "Teamworking", "On-the-Job Training", "No. Benefits", and "Monitoring" are provided on an ordinal scale which leads us to use an ordered probit estimation approach.

The first set of controls, $\mathbf{I_{id}}$, comprises competency tests and a compound measure, whether employees (i.e. non-managers) either receive performance payments or profit payments. We define this measure as "Incentive Pay". These two variables can (along with personality tests) be regarded as indicators for modern human resource practises, which itself may be correlated with all outcome variables we observe. Controlling for them, we hope to reduce the problem of omitted variables.[25]

Firm fixed controls are summarised in $\mathbf{X_i}$, containing dummies for all nine possible occupational groups in a firm. We include these dummies whenever running regressions on occupational group level. In all regressions, we control for whether a firm belongs to a foreign organisation or is unionised. Furthermore we control for detailed recruiting practises and account for region, industry, size of the establishment, and use a dummy which indicates whether the establishment belongs to the public sector. As explained

[24] The results are robust to the inclusion of managers. However, as the focus of this study is on reciprocal behaviour of employees, we exclude managers, who traditionally stand between the workforce and the owner of the company and hence may have different incentives.

[25] Note that we are fully aware of the difficulty to establish any causal effect of personality tests and we do not claim to be able to do so.

Table 2: Benefits for the Employee

	(1) OLS Bottom Wage	(2) OLS Top Wage	(3) O. Probit Training	(4) Probit Gen. Training	(5) Probit Job Security
Pers. Test	-0.051^{***}	-0.0066	0.26^{*}	0.34^{**}	0.22
	(0.014)	(0.015)	(0.14)	(0.17)	(0.14)
Comp. Test	-0.012	0.012	0.087	0.038	0.11
	(0.012)	(0.013)	(0.088)	(0.11)	(0.099)
Inc. Pay	-0.0059	0.049^{***}	0.14	0.13	0.042
	(0.014)	(0.012)	(0.097)	(0.12)	(0.12)
Foreign	-0.012	0.21^{***}	0.29	-0.29	-0.56
	(0.022)	(0.062)	(0.28)	(0.39)	(0.42)
Union	-0.034^{***}	-0.033^{**}	0.0061	-0.052	0.49^{***}
	(0.011)	(0.015)	(0.12)	(0.14)	(0.11)
PubSector	-0.043^{***}	0.00064	0.13	0.18	0.38^{**}
	(0.014)	(0.028)	(0.15)	(0.21)	(0.18)
Controls	Yes	Yes	Yes	Yes	Yes
Subpop. Observations	2115	2189	1888	1964	7892
R^2	0.170	0.395			

Notes: We report the coefficients and robust standard errors of OLS regressions of the share of employees within a firm, who earn less than 4.5 pounds per hour ("Bottom Wage", reported in column (1)), the share of employees earning the "Top Wage" (more than 15 pounds per hour, reported in column(2)) as well as of probit regressions of provision of "Training" (column (3)), "General Training" (column (4)), and "Job Security" (column(5)), on dummy variables personality tests, competency tests, and on controls. Regressions in the first two columns provide results on firm level, and estimates for column (3) and (4) report estimates for the largest occupational group. Column (5) provides estimates for each occupational group and includes an additional dummy to control for the number of occupational groups per firm.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in footnote 23, estimating the effect of personality tests on "Job Security" includes the number of occupational groups into firm-level controls.

3.3 Results

Table 2 and Table 3 summarise estimation results on Hypothesis 1, which we previously defined as necessary conditions in order to induce reciprocal behaviour of employees. We interpret personality tests as an indicator of whether employers search for potentially reciprocal workers. If they do so we should observe patterns associated with gift-exchange motivation.

Table 2 column (1) provides evidence that personality tests are significantly and nega-

tively related to the share of employees receiving very low wages of 4.5 pounds per hour or less. Note also, that this is true for personality tests, but not for the other two proxies for modern human resource practises, competency tests and variable payments. In contrast, we do not find that personality tests can explain the relative share of employees who earn top wages, see column (2). These results provide some tentative evidence for the presence of a gift-exchange motive when employers screen on personality traits upon hiring.

Firms which screen their job candidates for personality offer significantly more days of on-the-job training per year (column (3)) and are more likely to train their employees with general skills which are beneficial for their future working life (column (4)). Similarly to low wages, neither competency tests nor incentive payments predict the amount of on-the-job training. The same applies to the matters that the training covers: the coefficient on competency test for instance is ten times smaller than the estimate of personality tests.

Finally, there seems to be a weak tendency for firms with obligatory personality tests to provide more job security, as shown in column (5). However, this relationship is not significantly different from zero on any common level. In Section 3.4 we repeat our analysis with different sets of human resource controls. There we find persistent and significant correlations of job security with personality tests.

Table 3 exploits information about non-pay terms and conditions in more detail. In column (1) we find personality tests being associated with the likelihood that at least one of five suggested benefits is provided by the employer.[26] The intuition for this measure is that firms providing reciprocal incentives may face different costs for each of the listed benefits. Hence in order to make use of reciprocity most cost efficiently, different benefits may be chosen.

The next column provides evidence for a positive relation between personality tests and how many of five different benefits employees within a firm enjoy; this confirms the results from the first column. Finally, we analyse two non-pay terms: the presence of an employer pension scheme and the provision of more than four weeks annual paid leave are strongly correlated with the use of personality tests.[27]

These benefits are not only strongly associated with personality tests but similarly closely related to variable payments. Note, however, that competency tests only poorly predict the provision of these benefits. Only for employer pension schemes (column (3)) we find significant correlations for competency tests, confirming our assumption that

[26] The classification of these benefits are summarised in footnote 17.

[27] We report estimation results only for two out of five potential non-pay terms (c.f. footnote 17). These omitted conditions show systematically positive, though insignificant, correlations with personality tests. Estimation results are available from the authors upon request.

Table 3: Non-Pecuniary Benefits for the Employee

	Probit Benefits (1)	O. Probit No. Benefits (2)	Probit Pension (3)	Probit Extended Paid Leave (4)
Pers. Test	0.36*	0.29***	0.44***	0.36**
	(0.20)	(0.11)	(0.17)	(0.16)
Comp. Test	0.059	0.083	0.24**	0.13
	(0.13)	(0.078)	(0.11)	(0.11)
Inc. Pay	0.31**	0.36***	0.29***	0.34***
	(0.12)	(0.083)	(0.11)	(0.11)
Foreign	−0.35	0.0073	0.0029	−0.22
	(0.35)	(0.32)	(0.39)	(0.35)
Union	0.48**	0.39***	0.51***	0.52***
	(0.20)	(0.094)	(0.14)	(0.14)
PubSector	1.32***	0.43***	1.17***	0.83***
	(0.34)	(0.14)	(0.26)	(0.20)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275

Notes: We report the coefficients and robust standard errors of probit regressions of provision of "Benefits" (column (1)), provision of employer "Pension" scheme (column (3)), "Extended Paid Level" (column(4)) and ordered probit regressions of the "No. Benefits" (column (2)) on personality tests, competency tests, and on controls. All regressions provide estimates for the largest occupational group. Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

personality tests and competency tests are not substitutes.

Result 1 (Generosity to Workers) *Firms which screen for job candidates' personality are less likely to pay very low wages and provide more on-the-job training which then covers more general matters. Furthermore employees in firms with personality tests benefit from a higher likelihood to receive non-monetary benefits, especially employer pension schemes as well as extended paid leave and they receive a higher number of non-pay benefits overall.*

Rational employers provide gifts only if they expect to profit from this strategy. Hence the second set of hypotheses is concerned with benefits for the employers' side.

Table 4 summarises potential benefits for the employer. Column (1) reports correlation results of personality tests and team-working which are highly significant. Competency tests are associated with team-working as well, but comparing the magnitudes of the coefficients, it becomes clear that personality tests are associated with higher levels of

Table 4: Benefits for the Employer

	(1) O. Probit Teamworking	(2) O. Probit Monitoring	(3) OLS Dismissal	(4) Probit Performance	(5) Probit Firm Benefit	(6) Probit Firm Benefit 2
Pers. Test	0.43***	0.081	0.0045	0.26*	0.51***	0.28*
	(0.15)	(0.10)	(0.0059)	(0.14)	(0.14)	(0.15)
Comp. Test	0.17*	−0.035	−0.00075	−0.089	0.076	0.0012
	(0.095)	(0.083)	(0.0041)	(0.099)	(0.11)	(0.14)
Inc. Pay	0.039	0.067	0.0076*	0.19*	0.083	−0.18
	(0.092)	(0.089)	(0.0043)	(0.11)	(0.12)	(0.15)
Foreign	−0.20	−0.39	−0.018***	0.12	0.33	0.18
	(0.25)	(0.29)	(0.0060)	(0.36)	(0.33)	(0.37)
Union	−0.032	−0.00094	−0.0087**	−0.15	−0.17	0.100
	(0.12)	(0.12)	(0.0036)	(0.12)	(0.13)	(0.13)
PubSector	0.23	0.048	−0.0065*	−0.058	0.22	−0.31
	(0.18)	(0.16)	(0.0038)	(0.19)	(0.24)	(0.23)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2268	2279	2149	2147	2279	2279
R^2			0.075			

Notes: We report the coefficients and robust standard errors of ordered probit regressions of the degree of "Team-working" (column (1)), and "Monitoring" (column (2)), OLS regressions on the relative share of the variable "Dismissal" within one year in column (3) and probit regressions on firm "Performance" (column (4)) and in columns (5) and (6) two compound measures for overall "Firm Benefit" on dummy variables personality tests, competency tests, and on controls. Regression in the column (1) is based on the largest occupational group and column (2) – (6) provide results on firm level.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

team-working. Furthermore it is important to notice that incentive pay is uncorrelated with the likelihood of what fraction of employees are designated to teams.

The second two hypotheses are borrowed from Huang and Cappelli (2010). Contrary to their results, however, we do not find any relation between personality tests (or any other modern human resource practise) and monitoring. At first glance this is counterintuitive, as it seems to contradict reciprocity as underlying story of our results. However, this may be a result of the design of the question in the WERS, as it asks for the fraction of "non-managerial employees [who] have job duties *involving* supervising other employees" whereas Huang and Cappelli (2010) estimate the employee-supervisor ratio. Hence, in our data personality tests may reduce (unobserved) payrolls for employees who exclusively monitor as reciprocal employees control each other and do not collude against the employer. This implies that results may differ, as variables measure different dimensions of monitoring.

Second, there seems to be not a strong relation between relative dismissals and any of

the suggested human resource practises including personality tests. The absence of such a relation, however, questions the argument that firms use personality tests primarily to ensure an employee's "fit" to a company rather than also identifying specific desirable traits, like social preferences and reciprocity. If ensuring fit was the primary concern, we would expect to see *fewer dismissals* in firms that use personality tests. Thus we conjecture that personality tests are not merely used to improve general fit but may also be devices to screen for social preferences.

We address potential critique of limited variation in firms reporting dismissals by providing a mesaure which includes employees who left or resigned voluntarily. This behaviour may imply a rationale for workers to preempt dismissals. Once these cases are included, 15 percent of firms report turnover within the last year. Our result that personality tests are unrelated to turnover however is unaffected by this manipulation.

Column (4) provides evidence that firms using personality tests seem to perform better, at least according to self-rated performance measures. Similarly to previous results competency tests have no explanatory power, whereas incentive pay predicts success comparably well.

Finally, columns (5) and (6) report results of the constructed measure on firm's benefits. "Firm Benefit" – i.e. whether work flows are organised in teams, the firms uses little monitoring, or reports high productivity – is highly related to personality tests and screening for personality is the only dimension of modern human resource practises which has predictive power. Furthermore, the alternative measure "Firm Benefits 2", which additionally to "Firm Benefits" assigns benefits to the firm if not a single employee was dismissed within the preceding year provides similar evidence.[28] For these two measures it is most striking that solely personality tests can explain, whether firms are profiting in at least one of the suggested dimensions. However as personality tests and dismissals are unrelated, lower coefficients of "Firm Benefits 2" compared to "Firm Benefits" are not surprising.

Result 2 (Value to Firms) *Firms which screen for job candidates' personalities designate more employees to work in teams and report to be more successful on the market. Pooling potential benefits, more team-working, less monitoring, better market performance (and less dismissals in a second specification) are highly related to the use of personality tests in hiring.*

[28] See section 2 for the proper definition of "Firm Benefits 2".

3.4 Robustness

By providing robustness tests for the previous results, this section also offers an extensive discussion of our results so far. We are aware that drawing causal inferences is not valid as we cannot argue that personality tests were randomly assigned to firms. However, we go to great lengths to control for the general sophistication of a firm's human resource department. Modern human resource devices like personality tests or competency tests are likely to be correlated with (unobservable) other dimensions of quality of management practises which itself may be related to suggested benefits as well. Without being able to entirely exclude this mechanism, we aim to address that shortcoming by applying different sets of human resource practises as control variables.

We suggest five sets of human resource practises, only affecting the vector $\mathbf{I_{id}}$. Set 1 only includes whether the respective firm requires competency tests upon hiring and neglects incentive payments. Thus the indicator vector $\mathbf{I_i}$ only varies across firms, not within firms. The second set, Set 2, additionally includes whether the firm asks for personality tests for managers.[29] Set 3 additionally includes incentive payments. For the last two human resource sets we construct indicators reflecting potential complementarities between these measures: In Set 4 the indicator for modern human resource practises is equal to one if at least one of three, competency tests, personality tests *for managers* or incentive pay, is present at the respective firm. This measure has the least strict requirements for a firm to be classified as using modern human resource policies. In contrast to that, Set 5 requires firms to use all of the previously listed devices, implying it to be the strictest criterion for a classification in to the modern human resource category.

The appendix contains robustness tables on coefficients and standard errors for personality tests for each dependent variable and for each set of human resource controls. Note that even though not all coefficients of interest are significant at the highest level, the very systematic pattern of correlations emerging across a large set of specifications lends our core results substantial support. Table 5 to Table 9 refer to the table "Benefits for the Employee". Both bottom and top wages are summarised in Table 5 and Table 6. "Bottom Wage" is related to personality tests for each set of controls, whereas "Top Wage" is not correlated to personality tests across any specification.

We observe similar behaviour of personality tests on employee benefits. Personality tests are significantly associated with "On-the-Job Training" in three of five control sets (Table 7) and screening for personality of non-managers is significant in all regressions on "General Training" except when we explicitly include personality tests for managers

[29] Remember that managers are excluded in the entire analysis in order to avoid confounding results, as managers' job profiles involve both principal and agent duties.

(Table 8).

In Table 2, "Job Security" is positively though insignificantly related to personality tests. Regarding Table 9 we find significant associations in four of five specifications. With weak evidence from our main regressions, we conclude to only provide some tentative evidence in favour of higher job security in establishments with personality tests.

The next set of tables, Table 10 to Table 13 relate to dependent variables in Table 3. Common to all four tables is that in specification 1, 4 and 5 coefficients change to only minor degrees and standard errors are comparable. Control sets 2 and 3 on the contrary depict smaller impacts of personality tests on dependent variables which in most cases – with the exception of employer pension scheme – lead to insignificant coefficients of personality tests. However, explicitly including personality test for managers (the decisive criterion of Set 2 and Set 3) into our analysis of whether personality tests influence suggested benefits changes the situation: Performing an adjusted Wald test for joint significance of personality tests for managers and personality tests for non-managers provides evidence for joint importance of personality tests. Personality tests for managers and non-managers are jointly significant at the ten percent level in Set 2 for "Any Benefit" (Table 10). Both tests are jointly significant for both sets for "No. Benefits" on a one percent level, as can be seen in Table 11. Finally, both tests are jointly significant for "Pension Scheme" and "Extended Paid Leave" on a five percent level, as reported in tables 12 and 13.[30]

Finally, Tables 14 to Table 19 provide a closer look on all six regressions in Table 4, which summarises benefits for the employer. The correlation between personality tests of non-managers and team-working (Table 14) is stable and significant across all five specifications of modern human resource practises. Even with different sets of controls, hypotheses on less monitoring (Table 15) and reduced dismissals (Table 16) are unrelated to personality tests of non-managers. This is also true for joint significance for manager and non-manger screening of the establishment. These negative results on dismissals across all control sets provides further evidence that personality tests are not (only) applied to improve the "fit" between applicant and firm.

Table 17 shows significant correlations between personality tests of non-managers and "Firm Performance" for all but one control sets. The same applies to "Firm Benefit", the compound measure whether firms benefit at least in one dimension of less monitoring, more team-working, or better performance, as depicted in Table 18. Finally, there is a weaker relationship between screening for job candidates' personality and "Firm Benefit 2" (Table 19). This should not be surprising, as "Firm Benefit 2 " is defined as "Firm

[30]Details are available from the authors upon request.

Benefit" plus less dismissals. However, as shown previously, dismissals are not related to personality tests.

Summarising, Result 1 is robust to specifications 1, 4 and 5, but seems less robust regarding specifications 2 and 3, i.e. when personality tests for managers are explicitly included to the controls. However as many firms use personality tests for managers and non-managers provided that they use personality tests, both measures are highly correlated, resulting in imprecise point estimates. This is the reason why we reported adjusted Wald tests, which are by and large in line with the main regressions. Robustness tests for Result 2 do not systematically deviate from findings in the main section, suggesting that the association between personality tests and firm benefits seems to be profound.

4 Discussion

In previous years increasingly many contributions in personnel economics relate social preferences of employees to firm behaviour. Accounting for employees' (social) preferences may alter organisational structure within the firm and can lead to different job characteristics; see, e.g., Bartling et al. (2012).

In this paper we use the 2004 wave of the Workplace Employment Relations Survey (WERS 2004) and find that firms behave consistent with a model gift-exchange based motivation for their employees *if* they screen job candidates for personality. We use personality tests as a proxy for the degree of reciprocity (susceptibility to gift-exchange) within the workforce. Previous research has documented that traits elicited in personality tests are correlated with (laboratory) concepts of reciprocity.

Firms which apply personality tests are more likely to provide their employees possibly non-pecuniary benefits like employer pension schemes or grant extended paid annual leave. These employers are furthermore less likely to pay very low wages and provide more on-the-job training to their employees. The topics covered in the provided training are rather general instead of workplace related, implying a higher added value for workers. Finally, there is a weak tendency that firms with personality tests are more likely to provide their employees protection against redundancies via job security. On the other hand, firms also benefit from screening for personality: we find that these firms have higher rates of team-working and are generally more successful on the market.

Importantly, competency tests upon hiring and incentive pay, both modern human resource practises similarly to personality tests, predict only poorly (if at all) benefits both for the firm as well as for employees. This implies that the use of modern human resource practises is not sufficient to explain the provision of benefits and firm performance. It is

necessary that firms explicitly screen for job candidates' personality.

Closest to this study is Huang and Cappelli (2010). Using US survey data, they proxy for the importance of job candidates' "work ethic" for employers' hiring decisions. These authors find that firms which put high weight on "work ethic" on average pay higher wages, have more team-working and are more productive. Furthermore these firms monitor their employees to a lower degree and have fewer turnovers.

By and large, our analysis confirms the results of Huang and Cappelli (2010). Our results based on the WERS 2004 only deviate in two dimensions: First, we do not find stable relationships between screening and monitoring which however could be caused by different defintions on monitoring. Second, our measure of turnover – dismissals relative to firm size – is unrelated to personality tests. Note that the latter (negative) result also implies that personality tests are not primarily a device to improve the fit between applicant and firm. Together with poor predictive power of other human resource practises, gift-exchange motives in firms with personality tests seem to be a plausible explanation for our findings.

One aspect, that is present in the Englmaier and Leider (2012) discussion but was not highlighted here, lays out a trajectory for future research. To the extent that wage setting and benefit provision policies of firms are known to applicants, one would expect self-selection of workers into firms. We deem the analysis of the complementary use of hiring policies to attract the right talent and incentive provision and benefits design to motivate, develop, and retain this talent as one of the key challenges for personnel and labor economics.

References

Akerlof, G. A. (1982): "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, 97, 543–569.

Ashton, M., S. Paunonen, E. Helmes, and D. N. Jackson (1998): "Kin Altruism, Reciprocal Altruism, and the Big Five Personality Factors," *Evolution and Human Behavior*, 19, 243–255.

Autor, D. H. and D. Scarborough (2008): "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments," *Quarterly Journal of Economics*, 123, 219–277.

Bartling, B., E. Fehr, and K. Schmidt (2012): "Screening, competition, and job design: Economic origins of good jobs," *American Economic Review*, 102, 834–864.

Barton, H., J. 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, 465–497.

Becker, S. O., D. Messer, and S. C. Wolter (2013): "A Gift is Not Always a Gift: Heterogeneity and Long-term Effects in a Gift Exchange Experiment," *Economica*, 80, 345–371.

Ben-Ner, A., L. Putterman, F. Kong, and D. Magan (2004): "Reciprocity in a two-part dictator game," *Journal of Economic Behavior & Organization*, 53, 333–352.

Cabrales, A., R. Miniaci, M. Piovesan, and G. Ponti (2010): "Social Preferences and Strategic Uncertainty: An Experiment on Markets and Contracts," *American Economic Review*, 100, 2261–2278.

DellaVigna, S. (2009): "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 47, 315–372.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2009): "Homo Reciprocans: Survey Evidence on Behavioural Outcomes," *The Economic Journal*, 119, 592–612.

Dur, R. and A. Glazer (2007): "Optimal Contracts When a Worker Envies His Boss," *Journal of Law, Economics, and Organization*, 24, 120–137.

Englmaier, F. and S. Leider (2010): "Gift Exchange in the Lab - It is Not (Only) How Much You Give..." *CESifo Working Paper Series No. 2944*.

——— (2012): "Contractual and Organizational Structure with Reciprocal Agents," *American Economic Journal: Microeconomics*, 4, 146–183.

Englmaier, F., S. Strasser, and J. Winter (2014): "Worker Characteristics and Wage Differentials: Evidence from a Gift-Exchange Experiment," *Journal of Economic Behavior and Organizations*, 97, 185–203.

Englmaier, F. and A. Wambach (2010): "Optimal incentive contracts under inequity aversion," *Games and Economic Behavior*, 69, 312–328.

Fehr, E. and S. Gächter (2000): "Fairness and Retaliation: the Economics of Reciprocity," *Journal of Economic Perspectives*, 14, 159–181.

Fehr, E. and K. Schmidt (2003): "Theories of Fairness and Reciprocity - Evidence and Economic Applications," in *Dewatripont et. al. (Hg.) 2003 – Advances in Economics and Econometrics*.

Huang, F. and P. Cappelli (2010): "Applicant Screening and Performance-Related Outcomes," *American Economic Review*, 100, 214–218.

Ichniowski, C. and K. P. G. Shaw (1997): "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 87, 291–313.

Itoh, H. (2004): "Moral Hazard and Other-Regarding Preferences," *The Japanese Economic Review*, 55, 18–45.

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

Lesch, H. (2004): "Trade Union Density in Internatinal Comparison," *CESifo Forum*, 5, 12–18.

Leuven, E., H. van Oosterbeek, R. Sloof, and C. van Klavern (2005): "Worker Reciprocity and Employer Investment in Training," *Economica*, 72, 137–149.

Prendergast, C. (1999): "The Provision of Incentives in Firms," *Journal of Economic Literature*, 37, 7–63.

Rynes, S. and D. Cable (2003): "Recruitment Research in the Twenty-First Century," in *Handbook of Psychology*, 55–76.

Visser, J. (2006): "Union Membership Statistics in 24 Countries," *Monthly Labor Review*, 129, 38–49.

Wilk, S. and P. Cappelli (2003): "Understanding the Determinants of Employer Use of Selection Methods," *Personnel Psychology*, 56, 103–124.

Table Appendix

The following tables provide estimates for five different sets of modern human resource controls, other firm related controls are unchanged.[31] Set 1 only includes a dummy variable indicating whether the respective firm uses competency tests. On top of that, Set 2 controls for personality tests of managers, whereas Set 3 additionally includes incentive payments. Set 4 and Set 5 are compound measures for the presence of modern human resource practises: The dummy in Set 4 equals one if either the firm uses competency tests or personality tests for managers or incentive pay. The indicator in Set 5 is one if all suggested measures, competency tests, personality tests for managers and incentive pay are present at the firm.

Table 5: Robustness: Bottom Wage

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	−0.40**	−0.38*	−0.38*	−0.38*	−0.42**
	(0.19)	(0.21)	(0.20)	(0.20)	(0.19)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2115	2115	2115	2115	2115
Adj. Wald Test					
F(2, 2153)		2.17	2.13		
Prob > F		0.12	0.12		

Notes: We report the coefficients and robust standard errors of OLS regressions of the share of employees earning bottom wages (below 4.5 pounds per hour) on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

[31] We refer to the robustness section for an extensive discussion.

Table 6: Robustness: Top Wage

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.16	0.023	−0.014	0.13	0.17
	(0.15)	(0.18)	(0.18)	(0.15)	(0.17)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2113	2113	2113	2113	2113
Adj. Wald Test					
F(2, 2153)		3.06	2.60		
Prob > F		0.05	0.07		

Notes: We report the coefficients and robust standard errors of OLS regressions of the share of employees earning high wages (above 15 pounds per hour) on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 7: Robustness: Training

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.26*	0.25	0.25	0.25*	0.31**
	(0.14)	(0.15)	(0.15)	(0.13)	(0.14)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1888	1888	1888	1888	1888
Adj. Wald Test					
F(2, 2126)		1.95	1.90		
Prob > F		0.14	0.15		

Notes: We report the coefficients and robust standard errors of ordered probit regressions of how many days employees are trained during one year on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 8: Robustness: General Training

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.34*	0.23	0.23	0.32*	0.32*
	(0.17)	(0.19)	(0.19)	(0.17)	(0.18)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1964	1964	1964	1964	1964
Adj. Wald Test					
F(2, 2124)		3.36	3.22		
Prob > F		0.04	0.04		

Notes: We report the coefficients and robust standard errors of probit regressions of the provision of general training on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 9: Robustness: Job Security

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.28	0.33*	0.34*	0.32*	0.34*
	(0.18)	(0.20)	(0.20)	(0.18)	(0.18)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	6982	6982	6982	6982	6982
Adj. Wald Test					
F(2, 2165)		1.43	1.51		
Prob > F		0.24	0.22		

Notes: We report the coefficients and robust standard errors of probit regressions of the provision of job security on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of all occupational group and includes a control for the number of occupational groups per firm. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 10: Robustness: Benefits

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.37*	0.29	0.28	0.34*	0.43**
	(0.19)	(0.23)	(0.23)	(0.20)	(0.21)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275	2275
Adj. Wald Test					
F(2, 2187)		2.53	2.23		
Prob > F		0.08	0.11		

Notes: We report the coefficients and robust standard errors of probit regressions of the provision of benefits for the employees on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 11: Robustness: No. of Benefits

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.30***	0.21	0.21	0.28**	0.28**
	(0.11)	(0.15)	(0.15)	(0.12)	(0.12)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275	2275
Adj. Wald Test					
F(2, 2187)		5.30	5.22		
Prob > F		0.01	0.01		

Notes: We report the coefficients and robust standard errors of ordered probit regressions of the number of provided benefits for the employees on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 12: Robustness: Employer Pension Scheme

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.45***	0.38**	0.38**	0.47***	0.50***
	(0.17)	(0.19)	(0.19)	(0.17)	(0.18)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275	2275
Adj. Wald Test					
F(2, 2187)		3.47	3.44		
Prob > F		0.03	0.03		

Notes: We report the coefficients and robust standard errors of probit regressions of whether employer offer pension schemes on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 13: Robustness: Extended Paid Leave

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.38**	0.28	0.27	0.35**	0.43***
	(0.16)	(0.18)	(0.19)	(0.16)	(0.17)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2275	2275	2275	2275	2275
Adj. Wald Test					
F(2, 2187)		3.54	3.09		
Prob > F		0.03	0.05		

Notes: We report the coefficients and robust standard errors of probit regressions of whether employer offer extended paid leave on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 14: Robustness: Team-Working

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.43***	0.37**	0.37**	0.46***	0.48***
	(0.15)	(0.16)	(0.16)	(0.15)	(0.15)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2268	2268	2268	2268	2268
Adj. Wald Test					
F(2, 2187)		4.39	4.33		
Prob > F		0.01	0.01		

Notes: We report the coefficients and robust standard errors of ordered probit regressions of what share of employees is designated to teams on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the analysis of the largest occupational group. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 15: Robustness: Monitoring

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.086	0.11	0.10	0.081	0.082
	(0.10)	(0.12)	(0.12)	(0.10)	(0.11)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2279	2279	2279	2282	2282
Adj. Wald Test					
F(2, 2153)		0.39	0.36		
Prob > F		0.68	0.70		

Notes: We report the coefficients and robust standard errors of ordered probit regressions of the share of employees who have monitoring tasks on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 16: Robustness: Dismissals

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.0051	0.0024	0.0020	0.0039	0.0041
	(0.0059)	(0.0069)	(0.0069)	(0.0061)	(0.0062)
Firm Controls	Yes	Yes	Yes	Yes	Yes
Subpop. Observations	2149	2149	2149	2152	2152
Adj. Wald Test					
F(2, 2072)		0.67	0.54		
Prob > F		0.51	0.58		

Notes: We report the coefficients and robust standard errors of OLS regressions of the share of employees who have been dismissed during the previous year on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 17: Robustness: Firm Performance

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.28**	0.32*	0.31*	0.26*	0.23
	(0.14)	(0.17)	(0.17)	(0.14)	(0.15)
Firm Controls	Yes	Yes	Yes	Yes	Yes
Subpop. Observations	2148	2148	2148	2148	2148
Adj. Wald Test					
F(2, 2192)		1.98	1.79		
Prob > F		0.14	0.17		

Notes: We report the coefficients and robust standard errors of probit regressions of self-reported measure of firm performance being one if either managers report higher than median financial performance of their own firm, or higher labour productivity or higher product quality on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 18: Robustness: Firm Benefit

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.52***	0.63***	0.63***	0.52***	0.51***
	(0.14)	(0.16)	(0.16)	(0.14)	(0.14)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2279	2279	2279	2282	2282
Adj. Wald Test					
F(2, 2189)		8.54	8.45		
Prob > F		0.0002	0.0002		

Notes: We report the coefficients and robust standard errors of probit regressions of compound measure of firm benefit being one, if the firm either uses higher than median team-working, less than median monitoring or reports higher than median firm performance as defined in Table 17 on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 19: Robustness: Firm Benefit 2

	Control Set 1	Control Set 2	Control Set 3	Control Set 4	Control Set 5
Pers. Test	0.27*	0.14	0.14	0.31**	0.23
	(0.15)	(0.16)	(0.16)	(0.15)	(0.15)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2279	2279	2279	2282	2282
Adj. Wald Test					
F(2, 2189)		2.38	2.59		
Prob > F		0.09	0.08		

Notes: We report the coefficients and robust standard errors of probit regressions of compound measure of firm benefit being one as defined in Table 18 or has lower or equal to median turnover on personality tests and five different sets of controls. For further details on the control sets, see Appendix. All regressions are based on the firm level. The adjusted Wald test refers to the null hypothesis, that the coefficient of personality tests for managers and personality tests for non-managers are zero.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.