GOVERNANCE AND THE EFFICIENCY OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 504

Reciprocity in Organisations
Evidence from the UK

Florian Englmaier*
Thomas Kolaska**
Stephen Leider***

*/** University of Munich
*** University of Michigan Ross School of Business

March 2015

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Reciprocity in Organisations

Evidence from the UK*

Florian Englmaier Thomas Kolaska Stephen Leider †

March 17, 2015

Recent laboratory evidence suggests that personality traits, in particular social preferences, may affect contractual outcomes under moral hazard. Using the British Workplace Employment Relations Survey 2004 we find that behaviour of employers and employees is consistent with the presence of gift-exchange motives: firms that screen applicants for personality are less likely to pay low wages and more likely to provide (non-pecuniary) benefits. Firms likewise benefit from employee screening as they can implement more team-working and are generally more successful. Other human resource management practices 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 is a plausible explanation for our results.

Keywords: Reciprocity, Organisational Structure, Employee Compensation

JEL Classification: D22, M52

*We are grateful to Sascha Becker, Katharina Schüssler, the editor and two refrees for helpful comments and suggestions. 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.

†Englmaier: University of Munich, florian.englmaier@econ.lmu.de; Kolaska: University of Munich, thomas.kolaska@econ.lmu.de. Leider: University of Michigan Ross School of Business, leider@umich.edu.

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 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 prevalence of weak (monetary) incentives in real world contracts, while apparently still obtaining nontrivial effort, is not well explained by the standard approach in contract theory which emphasizes the use of output realizations as a (noisy) measure of employee effort to overcome the moral hazard problem.[3]

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.[4] 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 high-powered 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 provide 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 UK-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

[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] Note that models that emphasize multitasking in work tasks, following Holmström and Milgrom (1991), have been one explanation why we empirically observe rather muted explicit incentives. The use of career incentives is another potential explanation for the empirical observation; see, e.g. Lazear (1979) or Levin (2003).

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

traits that may be correlated with job candidates' inclination towards reciprocity. In line with gift-exchange motives, these firms are less likely to pay their employees particularly low wages and other non-pecuniary benefits like employer pension schemes and extended paid annual leave. 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 a key modern human resource management (HRM) practices and is considered to increase productivity by reaping synergies; see Barton et al. (2003). 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. Note, that the assumption is not that "firms benefit from teamwork under the condition of non-shirking", but that (some) firms may benefit from the implementation of teamwork, and that it is more likely that the implementation of teamwork is beneficial when workers are reciprocal.

In contrast, two additional measures for the presence of somewhat sophisticated HRM 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 HRM 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 HRM practices 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.[5]

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 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.[6] 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 studies 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 and we consider the additional evidence in favour of reciprocity as still only of suggestive nature.

Personality tests are only one potential dimension of how firms screen job candidates.[7] Other popular methods are interviews, reference letters and – widely used – competency tests.[8] Whereas the latter aim 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

[5]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.

[6]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.

[7]Wilk and Cappelli (2003) show that employers differ substantially in the way and the extent to which they make use of applicant screening.

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

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 correlated with economic preference measures, in particular measures of *reciprocity* as commonly defined in laboratory experiments. Ashton et al. (1998) conclude on basis of hypothetical questions that high "agreeableness" and high "emotional stability" are associated with high reciprocal altruism. Borghans et al. (2008) provide an extensive review of psychology personality traits and their relation to economic concepts. However, they cannot provide evidence on the relationship of social preferences and personality because there had been no studies examining the direct link prior to their study. Dohmen et al. (2008) study the link between personality, reciprocity, trust, and socio-demographic characteristics in the SOEP and find significant positive correlations between measures of positive reciprocity and *Big Five* traits, in particular conscientiousness and agreeableness, and negative correlations between measures of negative reciprocity and in particular agreeableness. Becker et al. (2012) directly study the link between psychological personality measures and economic preferences using both lab and survey data. Their results corroborate the findings in Dohmen et al. (2008). Ben-Ner et al. (2004) conduct a dictator game experiment in which recipients in an initial game become dictators in a second game. They show that – when the subjects remain within their pair – the amount sent back is strongly correlated with the amount received, a typical example of reciprocal behavior. They find that *Big Five* indicators "agreeableness" and "openness" are positively associated with the degree of reciprocal transfers; i.e., more agreeable or open individuals react more strongly to the dictator's decision in the initial game. Cognitive ability seems not to influence the propensity to reciprocate. Englmaier and Leider (2010) conduct a real-effort gift-exchange experiment in the laboratory and show that the *Big Five* indicator "agreeableness" positively predicts the extent of the effort response to a wage-gift. To sum up, also there are only few studies on this specific topic, the results indicate a positive relationship between psychological personality measures and positive reciprocity.

The psychological literature has collected substantial evidence that personality tests produce useful information to firms. In a seminal paper, Schmidt and Hunter (1998) conduct a meta-analysis on the predictive validity of different personnel selection criteria on future job performance of individuals to document that cognitive ability and "sample work" are the most important predictor of job performance, but personality measures also have high predictive validity. In the 1990s a number of studies and meta-analyses on the predictive validity of – in particular – the *Big Five* for future job performance have been conducted. Barrick et al. (2001) review the state of the literature on the relationship between personality and performance by conducting a

meta-analysis on meta-analyses and find a strong general relationship between conscientiousness and performance and more heterogeneous relationships between other *Big Five* dimensions and different criteria for performance (which depend – perhaps not surprisingly – on job characteristics, the sector, etc.).

In an economics study, Autor and Scarborough (2008) document the 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. Hence, the hiring method applied in this firm ceteris paribus should lead also to an above average reciprocal workforce.

The importance of social preferences for individual decisions has been documented in various studies; for an extensive survey see Fehr and Schmidt (2003), for field evidence DellaVigna (2009), and 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.[9] 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 HRM practises. Using firm data from steel finishing lines, Ichniowski et al. (1997) find that the use of HRM 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 lab-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 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

[9]See for references Itoh (2004), Dur and Glazer (2008) or Englmaier and Wambach (2010)

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 management and HRM practises, e.g., Bloom and VanReenen (2007), Ichniowski et al. (1997), cannot explain the whole set of observations and we feel confident that our results point to an additional explanatory role for a more nuanced, behavioural, interpretation 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. An Appendix and an Online Appendix collect tables for robustness analysis.

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.[10] 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.[11] 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.[12]

The WERS 2004, consisting of 2,295 establishments surveyed, is a representative sample of the British economy.[13] The number of employees per establishment varies

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

[11]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.

[12]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.

[13]"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.

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.[14] 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 manufacturing 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.[15]

Within each firm the WERS 2004 distinguishes between 9 different occupational groups.[16] 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 HRM 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 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 indi-

[14]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.

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

[16]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
Job Security	2295	0.17	0.37	0	0	0	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" and "Low Wage" are continuous fraction of dismissed employees, and employees with low earnings. "No. Benefits" counts the number of granted benefits, the remaining variables are binary.

cates 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.[17]

[17]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

This classification based on self-reporting, splits the data into two almost equal parts of rather successful and unsuccessful establishments.

The data contain rich information on various aspects of the workers' compensation and benefit package. "Job Security" is reported for each occupational group.[18] 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.[19] Finally, "Low Wage" measures the fraction of low-wage earners compared to all employees within a firm. The WERS defines the lowest wage category as wages equal 4.5 pounds per hour or below. The dataset provides three more wage categories: 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. "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

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.

[18] The explanation in the WERS handbook states: "Is there a policy of guaranteed job security or no-compulsory redundancies for any of these groups at this workplace?" In the UK, as opposed to many continental European countries, employment is in general less protected and closer to the employment-at-will model practiced in the US. Workplaces offering "job security" in general abstain from terminating an employee's contract for other than disciplinary reasons or in the case of a company undergoing restructuring.

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

can be beneficial for an employee's entire working life across different employers.[20]

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. HRM Practises

The WERS 2004 provides detailed information about HRM 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 somewhat sophisticated HRM practises.

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).[21] 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.[22] 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 supportive effects of market pressure on the introduction of HRM 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.

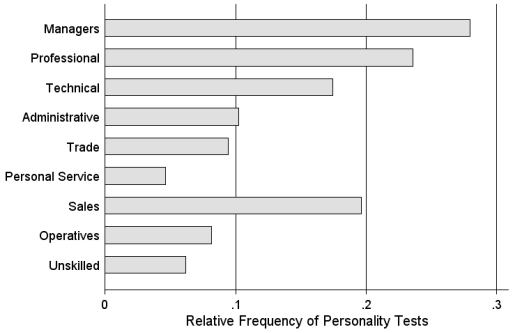
The prevalence of personality and competency tests by occupational group is summarised in detail in panel (a) and (b) in Figure 1. Comparing both panels it again becomes clear that employers use competency tests more often when hiring applicants: for each occupational group the relative frequency of competency tests exceeds that 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 occu-

[20] We present in Appendix B robustness checks on the definition of "General Training".

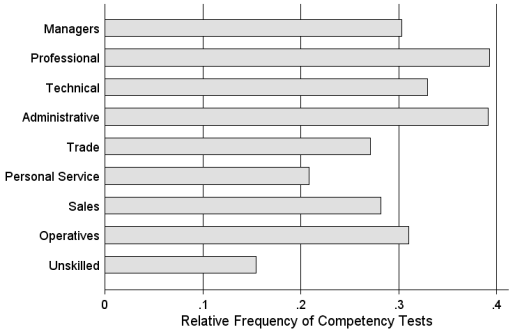
[21] The WERS handbook states that competency tests "target practical skills needed to perform the job, for example, a mechanic may be asked to dismantle and reconstruct a car engine."

[22] 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.

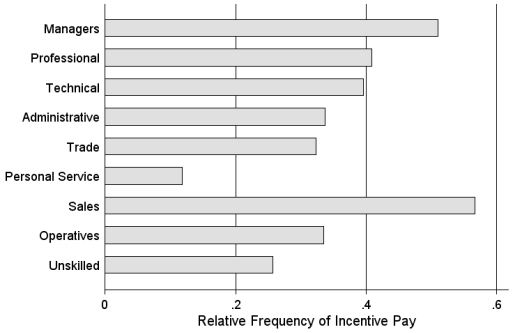
Figure 1: Relative Frequency of HRM Practises by Occupational Group
This figure provides an overview over the prevalence of personality tests, competency tests (upper row) and incentive pay for nine different occupational groups (bottom left). The picture on the bottom right depicts the prevalence of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



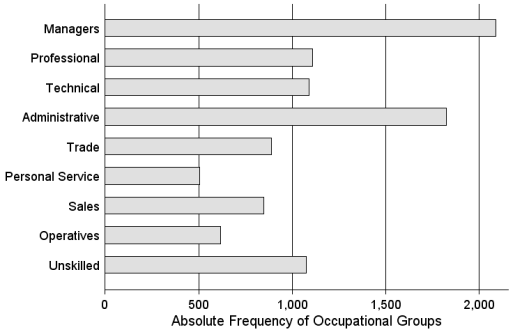
(a) Personality Tests



(b) Competency Tests



(c) Incentive Pay



(d) Occupational Group

pational 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.[23] 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.[24]

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

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

The third measure for advanced HRM 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).[25] 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.

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[26] 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

[25] 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.

[26] 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).

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 attachment 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 job security 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 illustrative 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 the 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.[27] $\mathbf{I_{id}}$ 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 at the 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 HRM practises and dependent variables on occupational group level. Second, different y_{id} are scaled differently, suggesting to adapt estimation strategies accordingly.

Job security – as defined in Footnote 19 – 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.[28] This implies that we are able

[27] In the main specification we present results where the indicator variable equals 1 if the firm uses personality tests in any occupational group. We re-run our analysis assigning value 1 for the personality test indicator only if the largest occupational group within a firm is screened using personality tests. All our results still hold under this specification. Results for this alternative specification are documented in Appendix E in Table E-15.

[28] 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

to match the provision of job security for each occupational group with the employed set of HRM 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-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 $\mathbf{I}_{\mathbf{i}\widetilde{\mathbf{d}}_\mathbf{i}}$. 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", and "Firm Performance", 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.[29] 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", and "Dismissal" are continuous variables, suggesting OLS estimation, for "Low Wages" we additionally estimate a Poisson model to account for the large number of observations at or close to zero. 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", and "Firm Performance". Finally "Team-working", "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 somewhat sophisticated HRM practises, which itself may be correlated with all outcome variables we observe.

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.

[29] 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.

Controlling for them, we attempt to mitigate the problem of omitted variables.

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 in Footnote 28, estimating the effect of personality tests on "Job Security" includes the number of occupational groups into firm-level controls.

Table 2: Benefits for the Employee

	(1) OLS Bottom Wage	(2) Poisson Bottom Wage	(3) O. Probit Training	(4) Probit Gen. Training	(5) Probit Job Security
Pers. Test	-0.051^{***}	-1.44^{***}	0.26^{*}	0.34^{**}	0.22
	(0.014)	(0.55)	(0.14)	(0.17)	(0.14)
Comp. Test	−0.012	−0.075	0.087	0.038	0.11
	(0.012)	(0.27)	(0.088)	(0.11)	(0.099)
Inc. Pay	−0.0059	0.025	0.14	0.13	0.042
	(0.014)	(0.25)	(0.097)	(0.12)	(0.12)
Foreign	−0.012	−2.01	0.29	−0.29	−0.56
	(0.022)	(1.32)	(0.28)	(0.39)	(0.42)
PubSector	-0.043^{***}	-1.49^{**}	0.13	0.18	0.38^{**}
	(0.014)	(0.66)	(0.15)	(0.21)	(0.18)
Union	-0.034^{***}	-0.82^{**}	0.0061	−0.052	0.49^{***}
	(0.011)	(0.36)	(0.12)	(0.14)	(0.11)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2104	2104	1888	1964	7892

Notes: We report the coefficients and robust standard errors of OLS and Poisson regressions of the fraction of workers with income less than 4.5 pounds/hour (columns (1) and (2)) as well as of (ordered) 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: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 a proxy for 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 negatively 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 that we employ to capture sophisticated HRM practises, competency tests and variable payments.[30] In order to substantiate the finding that firms using personality tests have fewer low wage earners, we amend our analysis and in column (2) also show results for a regression assuming a Poisson distribution – accounting for the large probability mass at and near zero – which confirms our findings from the OLS model.[31]

Firms which screen their job candidates for personality also 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)).[32] 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, for our main specification this relationship is not significantly different from zero at any common level. However, in Section 3.4 we repeat our analysis with varying sets of human resource controls and find persistent and significant correlations of job security with personality tests.[33]

[30] In principle, models of reciprocal motivation deliver relatively precise predictions on the optimal level and structure of wages. Unfortunately, the WERS has rather coarse information on these variables. In future work it would be highly desirable to be able to assess these predictions with more detailed data on wage levels and structures and at the same time on the composition of the workforce (to control for skill related compositional changes of the workforce in firms that use personality tests).

[31] In contrast, in unreported regressions we do not find significant evidence that personality tests can explain the relative share of employees who earn top wages.

[32] In Table B-2 we provide estimates for alternative definitions of our "General Training" variable: (1) we define "General Training" if the employee receives at least one of the following three trainings: Teamworking, Communicational Skills, or Leadership Skills; (2) includes Problem-Solving Skills; (3) includes in addition to (2) training on Reliability and Working towards Deadlines; (4) adds Quality Control Processes to (3); (5) uses Teamworking, Leadership Skills and Problem-Solving Methods to define "General Training". The results are robust to using either of these definitions.

[33] Note that for job security we have information on its provision across all occupational groups. Hence, Column (5) of Table 2 presents results of a regression where we use information on all occupational groups within a firm while we control for firm fixed effects; thus the substantially higher number of observations. I.e., we compare occupational groups with and without personality tests w.r.t. the incidence of job security provision.

Table 3: Non-Pecuniary Benefits for the Employee

	(1) Probit Benefits	(2) O. Probit No. Benefits	(3) Probit Pension	(4) Probit Extended Paid Leave
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 (ordered) 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.

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. 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 of employment: the presence of an employer pension scheme and the provision of more than four weeks annual paid leave are strongly and positively correlated with the use of personality tests.[34]

These benefits are not only strongly associated with personality tests but similarly

[34]We report estimation results only for two out of five potential non-pay terms of employment. The omitted terms of employment show systematically positive, though insignificant, correlations with personality tests. Estimation results are available from the authors upon request.

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 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 that is more general. Furthermore, employees in firms with personality tests are more likely to receive non-monetary benefits, especially employer pension schemes as well as extended paid leave. Such employees receive a higher number of non-pay benefits overall.*

Table 4: Benefits for the Employer

	(1) O. Probit Teamworking	(2) O. Probit Monitoring	(3) OLS Dismissal	(4) Probit Performance
Pers. Test	0.43***	0.081	0.0045	0.26*
	(0.15)	(0.10)	(0.0059)	(0.14)
Comp. Test	0.17*	−0.035	−0.00075	−0.089
	(0.095)	(0.083)	(0.0041)	(0.099)
Inc. Pay	0.039	0.067	0.0076*	0.19*
	(0.092)	(0.089)	(0.0043)	(0.11)
Foreign	−0.20	−0.39	−0.018***	0.12
	(0.25)	(0.29)	(0.0060)	(0.36)
Union	−0.032	−0.00094	−0.0087**	−0.15
	(0.12)	(0.12)	(0.0036)	(0.12)
PubSector	0.23	0.048	−0.0065*	−0.058
	(0.18)	(0.16)	(0.0038)	(0.19)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2268	2279	2149	2147
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 of the relative share of the variable "Dismissal" within one year in column (3) and Probit regressions of firm "Performance" (column (4)) on personality tests, competency tests, and on controls. Regression in the column (1) is based on the largest occupational group and column (2) – (4) provide results on firm level.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

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 the correlation between personality tests and team-working, which is 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 team-working. Furthermore it is important to notice that incentive pay is uncorrelated with the the fraction of employees that 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 HRM practise) and monitoring.[35] 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 not to be a strong relation between relative dismissals and any of the suggested HRM 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.

One potential concern is the limited variation in firms reporting dismissals. To address this we consider an alternate measure of employee turnover which includes employees who left or resigned voluntarily, which would include workers leaving to preempt dismissal. Once these cases are included, 15 percent of firms report turnover within the last year. However, this does not change our conclusion that personality tests are unrelated to turnover.

Finally, 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.

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.*

[35] Huang and Cappelli (2010) argue, plausibly, that being able to reduce the intensity of monitoring is beneficial for firms as it safes on costs.

In Appendix D we present evidence that the use of personality tests for hiring purposes is also highly related to various compound measures of firm benefits – pooling potential benefits like more team-working, less monitoring, better market performance and less dismissals in a second specification.

3.4. Robustness

By providing robustness tests for the previous results, this section also offers an extensive discussion of our results so far. A first obvious plausibility check is to see whether the use of personality tests in one occupational group relates to benefits for employees in another occupational group for which personality test are not used. We would expect that, if the effect is causal, the estimated coefficients are smaller in occupational groups in the same firms for which personality test are not used. Unfortunately, this check can only be performed for job security, the only item where we have information across all occupational groups within a firm.[36] To do so we include a dummy for occupational groups who have no personality test but belong to a firm that does use personality tests for other occupational groups. Inspecting the respective results in Table A-1 in Appendix A document that the coefficient for this dummy is in fact significantly lower than the coefficient for personality test.

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 HRM practises as control variables.

We suggest five sets of HRM practises to include in 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.[37] 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 sophisticated HRM 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

[36] We are grateful to two referees for suggesting this robustness check to us.

[37] 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.

requirements for a firm to be classified as using sophisticated HRM practices. 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 sophisticated HRM-use category.

The online appendix contains robustness tables with coefficients and standard errors for personality tests, competency tests, and for the two presented compound measures of HR sophistication 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 C-1 to Table C-5 refer to the table "Benefits for the Employee". "Low Wages" are summarised in Table C-1 and Table C-2. "Low Wage" is related to personality tests for each set of controls.

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 C-3) and screening for personality of non-managers is significant in all regressions on "General Training" except when we explicitly include personality tests for managers (Table C-4).

In Table 2, "Job Security" is positively though insignificantly related to personality tests. Regarding Table C-5 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 C-6 to Table C-9 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 "Benefit" (Table C-6). Both tests are jointly significant for both sets for "No. Benefits" on a one percent level, as can be seen in Table C-7. Finally, both tests are jointly significant for "Pension Scheme" and "Extended Paid Leave" on a five percent level, as reported in Table C-8 and Table C-9.[38]

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

Finally, Table C-10 to Table C-13 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 C-10) is stable and significant across all five specifications of HRM practises. Even with different sets of controls, hypotheses on less monitoring (Table C-11) and reduced dismissals (Table C-12) 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.

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, conditional on using personality tests at all, many firms choose to use personality tests for both managers and non-managers, hence both measures are highly correlated resulting in imprecise point estimates. For this reason we report 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. Conclusion

A large and growing body of recent research in personnel economics relates the 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 of 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 (non-pecuniary) benefits such as employer pension schemes or 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. On-the-job training is more frequent, and more general in content, rather than workplace related, implying a higher added value for workers. Finally, there is a weak tendency for firms with personality tests 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 are generally more successful (in terms of profits, productivity and/or quality) and are more likely to be able to employ team-working. Importantly, competency tests upon hiring and incentive pay predict only poorly (if at all) benefits both for the firm as well as for employees. This indicates that the use of sophisticated HRM practises is not sufficient to explain the provision of benefits and firm performance. It is necessary that firms explicitly screen for job candidates' personality.

By and large, our analysis confirms the results of Huang and Cappelli (2010). Using US survey data, they proxy for the importance of job candidates' "work ethic" for employers' hiring decisions and find that firms which put high weight on "work ethic" on average pay higher wages, have more team-working and are more productive. Our results only deviate in two dimensions: First, we do not find stable relationships between screening and monitoring – however this could be caused by different defintions of monitoring. Second, our measure of turnover – dismissals relative to firm size – is unrelated to personality tests. The latter (negative) result indicates that personality tests are not primarily a device to improve the fit between applicant and firm. Together with poor predictive power of other HRM practises, gift-exchange motives in firms with personality tests appear to have additional explanatory power for our findings.

In terms of future research we believe two avenues to be particularly fruitful. First, to the extent that the wage setting and benefit provision policies of firms are known to applicants, one would expect self-selection of workers into firms. Hence the analysis of the complementary use of hiring policies to attract the right talent, and the provision of incentives and benefits designed to motivate, develop, and retain this talent appears to be important.[39] Second, as stated in the introduction, incentive provision via career concerns and relational contracts is an important aspect of incentives in labour markets. Studying the interaction of dynamic incentives and reciprocity based incentives is certainly an important field of future study.[40]

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.

[39]See Englmaier and Schüssler (2015) for a more detailed research agenda along these lines.

[40]Dur and Tichem (forthcoming) analyses theoretically the viability of relational incentive contracts in the context of distributional, but not reciprocal, preferences.

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

Barrick, M. R., M. K. Mount, and T. A. Judge (2001): "Personality and performance at the beginning of the new millennium: What do we know and where do we go next?" *International Journal of Selection and Assessment*, 9, 9–30.

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, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012): "The Relationship Between Economic Preferences and Psychological Personality Measures," *Annual Review of Economics*, 4, 453–478.

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.

Bloom, N. and J. VanReenen (2007): "Measuring and Explaining Management Practices Across Firms and Countries," *Quarterly Journal of Economics*, 122, 1351–1408.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Ter Weel (2008): "The economics and psychology of personality traits," *Journal of Human Resources*, 43, 972–1059.

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 (2008): "Representative trust and reciprocity: prevalence and determinants," *Economic Inquiry*, 46, 84–90.

——— (2009): "Homo Reciprocans: Survey Evidence on Behavioural Outcomes," *The Economic Journal*, 119, 592–612.

Dur, R. and A. Glazer (2008): "Optimal Contracts When a Worker Envies His Boss," *Journal of Law, Economics, and Organization*, 24, 120–137.

Dur, R. and J. Tichem (forthcoming): "Altruism and Relational Incentives in the Workplace," *Journal of Economics and Management Strategy*.

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. and K. Schüssler (2015): "Complementarities of HRM Practices - A Case for Employing Multiple Methods and Integrating Multiple Fields," *CESifo Working Paper No. 5249*.

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*.

Holmström, B. and P. Milgrom (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Owenership, and Job Design," *Journal of Law, Economics, & Organisation*, 7, 24–52.

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

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, 291–313.

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

Lazear, E. P. (1979): "Why Is There Mandatory Retirement?" *Journal of Political Economy*, 87, 1261–1284.

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 Klaveren (2005): "Worker Reciprocity and Employer Investment in Training," *Economica*, 72, 137–149.

Levin, J. (2003): "Relational Incentive Contracts," *American Economic Review*, Vol. 93, 835–857.

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.

Schmidt, F. L. and J. E. Hunter (1998): "The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings," *Psychological Bulletin*, 124, 262–274.

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.

A. Robustness Analysis: Personality Tests in other Occupational Groups

Table A-1: Robustness: Personality Tests in other Occupational Groups

	(1) OLS Bottom Wage	(2) Poisson Bottom Wage	(3) O. Probit Training	(4) Probit Gen. Training	(5) Probit Job Security
Pers. Test	−0.051***	−1.44***	0.26*	0.34**	0.19
	(0.014)	(0.55)	(0.14)	(0.17)	(0.15)
Comp. Test	−0.012	−0.075	0.087	0.038	0.13
	(0.012)	(0.27)	(0.088)	(0.11)	(0.100)
Inc. Pay	−0.0059	0.025	0.14	0.13	0.047
	(0.014)	(0.25)	(0.097)	(0.12)	(0.12)
Foreign	−0.012	−2.01	0.29	−0.29	−0.57
	(0.022)	(1.32)	(0.28)	(0.39)	(0.42)
Pub. Sector	−0.043***	−1.49**	0.13	0.18	0.38**
	(0.014)	(0.66)	(0.15)	(0.21)	(0.18)
Union	−0.034***	−0.82**	0.0061	−0.052	0.49***
	(0.011)	(0.36)	(0.12)	(0.14)	(0.11)
Pers. Test (other)					−0.27*
					(0.15)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2104	2104	1888	1964	7892
Adj. Wald Test					
(1) Pers. Test = 0					
(2) Pers. Test (other) = 0					
F(2, 2185)					4.11
Prob > F					0.0165

Notes: This table reproduces Table 2 but includes the dummy variable "Pers. Test (other)" in column (5). This dummy is 1, if no personality tests are conducted in the occupational group at hand and at the same time in at least one other occupational group within the same firm personality tests are performed and 0 otherwise.
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

B. Robustness Analysis: Definition of General Training

Table B-2: Robustness: Variable Definitions of General Training

	(1) Probit Base Regression	(2) Probit Set 1	(3) Probit Set 2	(4) Probit Set 3	(5) Probit Set 4
Pers. Test	0.34**	0.32*	0.41**	0.61***	0.31*
	(0.17)	(0.17)	(0.18)	(0.20)	(0.16)
Comp. Test	0.038	0.094	0.075	0.046	0.15
	(0.11)	(0.11)	(0.11)	(0.12)	(0.11)
Inc. Pay	0.13	0.16	0.19	0.28**	0.16
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
Foreign	−0.29	−0.36	−0.35	−0.20	−0.29
	(0.39)	(0.38)	(0.38)	(0.39)	(0.39)
Union	−0.052	−0.076	−0.12	−0.063	0.020
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
PubSector	0.18	0.14	0.091	−0.025	0.057
	(0.21)	(0.21)	(0.21)	(0.22)	(0.20)
Constant	−0.37	−0.42	−0.33	−0.29	−0.71*
	(0.41)	(0.41)	(0.42)	(0.43)	(0.40)
Controls	Yes	Yes	Yes	Yes	Yes
Subpop. Observations	1964	1964	1964	1964	1964

Notes: We report the coefficients and robust standard errors of Probit regressions for different specifications of the variable "General Training". Column (1) shows the base regression from Table 2 where we define "General Training" if the employee receives at least one of the following three trainings: Teamworking, Communicational Skills, or Leadership Skills. Column (2) includes Problem-Solving Skills, column (3) includes (in addition to column (2)) training on Reliability and Working towards Deadlines. Column (4) adds Quality Control Processes to column (3) and column (5) uses Teamworking, Leadership Skills and Problem-Solving Methods to define "General Training". Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendix:

Reciprocity in Organisations - Evidence from the UK

by
Florian Englmaier & Thomas Kolaska &
Stephen Leider

C. Robustness Analysis: Differing Set of Controls

The following tables provide estimates for five different sets of human resource controls:

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 HRM 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.

Other firm related controls are unchanged: In all regressions, we control for whether a firm belongs to a foreign organisation or is unionised. Furthermore we control for region, industry, size of the establishment, and use a dummy which indicates whether the establishment belongs to the public sector.

Table C-1: Robustness: Low Wage – OLS

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	-0.052^{***}	-0.053^{***}	-0.053^{***}	-0.051^{***}	-0.054^{***}
	(0.014)	(0.015)	(0.015)	(0.013)	(0.014)
Comp. Test	−0.012	−0.012	−0.012		
	(0.012)	(0.012)	(0.012)		
Any HR				−0.026	
				(0.016)	
Full HR					−0.000078
					(0.014)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2100	2100	2100	2100	2100
Adj. Wald Test					
F(2, 2138)		7.46	7.21		
Prob > F		0.001	0.001		

Notes: We report the coefficients and robust standard errors of OLS regressions of the share of employees earning low 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 B. 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 C-2: Robustness: Low Wage – Poisson

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	−1.44***	−1.35**	−1.36**	−1.41***	−1.45***
	(0.55)	(0.59)	(0.59)	(0.54)	(0.56)
Comp. Test	−0.072	−0.067	−0.070		
	(0.28)	(0.28)	(0.27)		
Any HR				−0.22	
				(0.27)	
Full HR					−0.025
					(0.60)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2100	2100	2100	2103	2103
Adj. Wald Test					
(1) Pers. Test = 0					
(2) Groups Dummy = 0 F(2, 2185)				4.11	
Prob > F				0.0165	

Notes: We report the coefficients and robust standard errors of Poisson regressions of the share of employees earning low 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 B. 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 C-3: Robustness: Training

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.26*	0.25	0.25	0.25*	0.31**
	(0.14)	(0.15)	(0.15)	(0.13)	(0.14)
Comp. Test	0.096	0.093	0.086		
	(0.089)	(0.089)	(0.088)		
Any HR				0.25**	
				(0.11)	
Full HR					−0.16
					(0.17)
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 B. 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 C-4: Robustness: General Training

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.34*	0.23	0.23	0.32*	0.32*
	(0.17)	(0.19)	(0.19)	(0.17)	(0.18)
Comp. Test	0.045	0.028	0.022		
	(0.11)	(0.11)	(0.11)		
Any HR				0.21*	
				(0.13)	
Full HR					0.22
					(0.24)
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 B. 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 C-5: Robustness: Job Security

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.28	0.33*	0.34*	0.32*	0.34*
	(0.18)	(0.20)	(0.20)	(0.18)	(0.18)
Comp. Test	0.0024	0.011	0.018		
	(0.13)	(0.13)	(0.13)		
Any HR				−0.15	
				(0.16)	
Full HR					−0.31
					(0.21)
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 C-6: Robustness: Benefits

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.37*	0.29	0.28	0.34*	0.43**
	(0.19)	(0.23)	(0.23)	(0.20)	(0.21)
Comp. Test	0.084	0.074	0.049		
	(0.13)	(0.13)	(0.13)		
Any HR				0.24*	
				(0.13)	
Full HR					−0.15
					(0.29)
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 B. 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 C-7: Robustness: No. of Benefits

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.30***	0.21	0.21	0.28**	0.28**
	(0.11)	(0.15)	(0.15)	(0.12)	(0.12)
Comp. Test	0.10	0.093	0.073		
	(0.080)	(0.081)	(0.079)		
Any HR				0.23***	
				(0.085)	
Full HR					0.22
					(0.20)
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 B. 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 C-8: Robustness: Employer Pension Scheme

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.45***	0.38**	0.38**	0.47***	0.50***
	(0.17)	(0.19)	(0.19)	(0.17)	(0.18)
Comp. Test	0.26**	0.25**	0.23**		
	(0.11)	(0.11)	(0.11)		
Any HR				0.24**	
				(0.12)	
Full HR					0.092
					(0.29)
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 B. 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 C-9: Robustness: Extended Paid Leave

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.38**	0.28	0.27	0.35**	0.43***
	(0.16)	(0.18)	(0.19)	(0.16)	(0.17)
Comp. Test	0.16	0.15	0.12		
	(0.11)	(0.11)	(0.11)		
Any HR				0.31***	
				(0.11)	
Full HR					−0.031
					(0.22)
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 B. 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 C-10: Robustness: Team-Working

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.43***	0.37**	0.37**	0.46***	0.48***
	(0.15)	(0.16)	(0.16)	(0.15)	(0.15)
Comp. Test	0.18*	0.17*	0.16*		
	(0.095)	(0.096)	(0.095)		
Any HR				0.10	
				(0.10)	
Full HR					−0.019
					(0.19)
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 B. 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 C-11: Robustness: Monitoring

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.086	0.11	0.10	0.081	0.082
	(0.10)	(0.12)	(0.12)	(0.10)	(0.11)
Comp. Test	−0.034	−0.031	−0.032		
	(0.083)	(0.084)	(0.084)		
Any HR				−0.014	
				(0.093)	
Full HR					−0.013
					(0.14)
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 B. 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 C-12: Robustness: Dismissals

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.0051	0.0024	0.0020	0.0039	0.0041
	(0.0059)	(0.0069)	(0.0069)	(0.0061)	(0.0062)
Comp. Test	−0.00058	−0.00091	−0.0011		
	(0.0041)	(0.0042)	(0.0042)		
Any HR				0.0087**	
				(0.0044)	
Full HR					0.0042
					(0.0084)
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 B. 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 C-13: Robustness: Firm Performance

	Control Sets				
	(1)	(2)	(3)	(4)	(5)
Pers. Test	0.28**	0.32*	0.31*	0.26*	0.23
	(0.14)	(0.17)	(0.17)	(0.14)	(0.15)
Comp. Test	−0.084	−0.079	−0.083		
	(0.100)	(0.10)	(0.100)		
Any HR				0.032	
				(0.12)	
Full HR					0.14
					(0.21)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2147	2147	2147	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 than median labour productivity or higher than median product quality on personality tests and five different sets of controls. For further details on the control sets, see Appendix B. 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.

D. Compound Measure for Firm Benefits

We construct the variable "Benefit 1" 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, "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.

Variant B, for both compound measures, gives an alternative measure for "Firm Performance". Instead of including financial performance, labour productivity and superior product quality, "Firm Performance" in Variant B only then indicates better than market performance if the respective firm reports higher than median financial performance or labour productivity.

Table D-14 shows that both measures, "Benefit 1" and "Benefit 2", are strongly correlated with Personality Tests. This is true regardless of the assumed definition of "Firm Performance", i.e. for Variant B as well. It is most striking that only personality tests but not competency tests can predict, whether firms are profiting in at least one of the suggested dimensions.

Table D-14: Robustness: Firm Benefits – Compound Measure

	(1) Probit	(2) Probit	(3) Probit	(4) Probit
	Benefit 1		Benefit 2	
		Variant B		Variant B
Pers. Test	0.51***	0.54***	0.28*	0.33**
	(0.14)	(0.14)	(0.15)	(0.15)
Comp. Test	0.076	0.11	0.0012	−0.0047
	(0.11)	(0.11)	(0.14)	(0.13)
Inc. Pay	0.083	0.080	−0.18	−0.17
	(0.12)	(0.12)	(0.15)	(0.15)
Foreign	0.33	0.39	0.18	0.21
	(0.33)	(0.34)	(0.37)	(0.37)
Union	−0.17	−0.19	0.100	0.055
	(0.13)	(0.13)	(0.13)	(0.12)
PubSector	0.22	0.18	−0.31	−0.19
	(0.24)	(0.23)	(0.23)	(0.22)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2279	2279	2279	2279

Notes: We report the coefficients and robust standard errors of Probit regressions for different compound measures on firm benefits on Personality Tests, further HR practises and controls. The dependent variable in column (1) is 1 if the respective firm uses less monitoring than the median firm, or uses more teamworking than the median, or reports higher than median values for financial performance, labour productivity or product quality. The definition of the dependent variable in column (2) equals the definition in (1) excluding product quality as criterion. The indicator in column (3) uses the same definition as column (1) and additionally defines the dependent variable as success if the respective firm records zero dismissals within the previous year. The definition of the dependent variable in column (4) equals the definition in (3) excluding product quality as criterion. Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E. Robustness Analysis: Alternative Definition of the Use of Personality Tests

This table summarises outcome variables that are reported on the firm level. In contrast to previous regressions we define personality tests (competency tests/incentive pay) to be present if these measures are introduced in the largest occupational group instead of demanding the introduction of these measures in any occupational group.

As the results show, results are qualitatively not affected by this alternative definition.

Table E-15: Robustness: Alternative Definition of Personality/Competency Tests and Incentive Pay

	(1) OLS Low Wage	(2) Poisson Low Wage	(3) O. Probit Monitoring	(4) OLS Dismissal	(5) Probit Performance
Pers. Test	-0.059^{***}	-1.50^{**}	0.037	0.0060	0.27^{*}
	(0.016)	(0.59)	(0.11)	(0.0065)	(0.15)
Comp. Test	-0.0062	-0.091	-0.037	-0.000029	-0.061
	(0.013)	(0.30)	(0.087)	(0.0041)	(0.10)
Inc. Pay	0.0043	0.16	0.099	0.0080^{*}	0.20^{*}
	(0.015)	(0.25)	(0.092)	(0.0047)	(0.11)
Foreign	-0.014	-1.99	-0.40	-0.018^{***}	0.085
	(0.023)	(1.31)	(0.30)	(0.0060)	(0.37)
Union	-0.035^{***}	-0.84^{**}	0.0099	-0.0091^{**}	-0.16
	(0.011)	(0.37)	(0.12)	(0.0036)	(0.12)
PubSector	-0.043^{***}	-1.47^{**}	0.053	-0.0067^{*}	-0.064
	(0.013)	(0.67)	(0.16)	(0.0039)	(0.19)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	2105	2105	2284	2154	2149

Notes: This table summarises outcome variables that are reported on firm level. We define personality tests (competency tests/incentive pay) to be present if these measures are introduced in the largest occupational group instead of demanding the introduction of these measures in any occupational group. We report the coefficients and robust standard errors of OLS and Poisson regressions of the fraction of workers with income less than 4.5 pounds/hour (columns (1) and (2)) as well as of ordered Probit regressions of provision of "Monitoring" (column (3)), OLS regression of "Dismissal" (column (4)), and Probit regression of "Performance" (column(5)), on dummy variables personality tests, competency tests, and on controls.
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.