Good Jobs, Screening, and Labour Productivity

Evidence from the Field

Thomas Kolaska*

11th February 2014

Accumulating evidence shows that firm success is closely related to choosing the "right" mix of human resource management practises. This paper uses survey data from a representative sample of the British economy (WERS 2004) to identify clusters of policies which predict labour productivity. First, firms which leave their employees substantial discretion in working processes report (weakly) higher labour productivity despite increased shirking opportunities. Second, firms which pay high income to employees in addition to allowing discretion ("good jobs") are highly successful *if* they screen job candidates for personality. This result is consistent with gift-exchange behaviour between employers who pay high wages and employees who voluntarily exert high effort, as screening for personality predicts reciprocal behaviour. Third firms offering "good jobs" but abstain from screening (efficiency wage hypothesis) or rely on agent's "work ethic" without providing high wages (Huang and Cappelli (2010)) report significantly lower labour productivity than firms offering "good" jobs and screen for personality. Forth, tests aiming to uncover job candidate's ability in combination with "good" jobs predict beneficial firm outcomes substantially worse, implying that not screening itself but screening for personality is decisive for offering "good" jobs to predict high labour productivity.

Keywords: Organisational Structure, Human Resource Management Practises, Reciprocity
JEL Classification: D22, M54

*Department of Economics, University of Munich, Kaulbachstrasse 45, 80539 Munich, Germany; thomas.kolaska@econ.lmu.de. For useful comments, I thank Sascha Becker, Florian Englmaier, and Uwe Sunde.

1. Introduction

What is the optimal level of discretion employers should grant their workforce? On the one hand leaving employees a lot of leeway can increase productivity, as workers themselves might know best how to get their tasks done most efficiently.[1] This result has extensively been reviewed in Ichniowski and Shaw (2003). However, defining operating processes less strictly may induce employees to lower effort without having to fear immediate consequences. The latter argument is in line with classical moral hazard frameworks where agents maximise utility by minimising effort costs – at the expense of the principal who may for several reasons not be able or willing to fully monitor employees' actions. This basic trade-off between monitoring and shirking gave rise to fruitful research on incentive mechanisms which aim to circumvent one of the two extreme outcomes: either literally no discretion at the cost of high efficiency losses or some amount of discretion combined with high levels of shirking. To overcome that dilemma, standard theory usually suggests to pay agents for performance or, alternatively, provide efficiency wages to avoid potentially high risk premia and to cope with situations in which monitoring is particularly difficult.[2] An efficiency wage strategy is promising when both parties repeatedly interact such that the employee has a high continuation value from the relationship and the threat of dismissal is credible.

More recently, increasingly many contributions highlight the importance of non-standard behaviour of employees. Among others, two distinct mechanisms may both lead to employees providing higher effort even in situations with moral hazard: Huang and Cappelli (2010) find that self-reported employee performance seems to be higher if managers prefer to employ job candidates with high "work ethic". It is argued that these employees unconditionally exert higher levels of effort as working hard is part of their personality. This trait makes them more desirable for firms to hire.

Great attention in recent research however was also dedicated to the concept of reciprocity which implies gift-exchange behaviour between employer and employee.[3] Employees with reciprocal traits respond to kind labour contracts – commonly achieved with higher than outside option income for them – with high levels of effort. Compared to the concept of efficiency wages, which is agnostic about (non-standard) agent behaviour and to the mechanism of "work ethic" which assumes employees to intrinsically work harder, firms attempting to make use of reciprocity need to meet two conditions: in a world with heterogeneous agents, the firm first has to employ workers with reciprocal traits

[1] The terms "worker" and "employee" are used interchangeably throughout this essay.
[2] See Prendergast (1999) for an extensive survey on incentive mechanisms.
[3] For comprehensive surveys on reciprocity, see Fehr and Gächter (2000) or Fehr and Schmidt (2006).

and second has to offer these employees generous labour contracts. The first part can be achieved by explicit screening for reciprocity whereas the second part is most easily accomplished by paying higher than outside option incomes.

In this paper I use data from the 2004 wave of the "Workplace Employment Relations Survey" (WERS 2004), a representative survey of the British economy, to explore complementarities between human resource management practises (HRM) and labour productivity. Common sense suggests that firms deciding to grant employees high discretion may need to implement complementary practises to prevent workers from exploiting increased shirking opportunities. Contrary to this intuition I will find that firms leaving discretion do not face detrimental consequences even if discretion is not complemented with additional HRM practises which suggests that shirking may (here) not be the predominant effect. However combining high discretion with providing high income (this combination is henceforth referred to "good jobs", a terminology which is borrowed from Bartling et al. (2012a)) along with screening job candidates for personality significantly predicts higher than average labour productivity. Importantly, the necessary combination of personality tests upon hiring *and* high income suggests a conditional relationship which is consistent with gift-exchange as an underlying mechanism: generous firms may only expect high effort exertion (and hence high labour productivity) by the "right" employees, i.e. those who were screened for personality and hence should exhibit reciprocal traits. Simply providing "good jobs" alone without screening for personality is not sufficient to predict high firm performance. This study therefore casts doubts on efficiency wages as the optimal strategy to efficiently make use of discretion. Complementing discretion with personality tests for job candidates in hope to select the "right" employees who unconditionally exert high effort even for low wages likewise fails to be related to beneficial firm outcomes. Finally competency tests – a screening device to uncover applicant's cognitive ability – is not suitable to predict labour productivity to a similar extent as personality tests do.[4] This suggests that not screening itself but screening for personality is the key parameter to complement "good jobs", as only personality tests reveal reciprocal types.

Relating firm outcomes to generosity in income and discretion requires me to match responses from two sources of the WERS 2004 dataset. Labour productivity is taken from the management survey, which supplies firm level estimates of labour productivity. Estimates for the level of discretion and the generosity of income are derived from an

[4]Screening usually can be distinguished in two main dimensions: screening for personality and screening for ability/competency. Interviews, computer-based tests, and reference letters are popular screening devices and are applied to both dimensions.

employee questionnaire, for which a number of employees is drawn randomly from each firm surveyed in the WERS 2004. As income and the degree of discretion depend on several personal characteristics of the respective employee, I separately regress income and discretion on a series of observables like age, tenure, or occupational group to generate estimated values of both variables.[5] I then compare estimated and actual values of income and discretion separately to determine the degree of firm's generosity towards each single employee for each dimension. As the dataset does not provide information on individual performance or effort exertion I aggregate employee responses on firm level by calculating average scores of deviations of income and discretion for each establishment.

With personality tests and competency tests I use two distinct screening procedures job candidates may have to undergo. Though both practises are commonly used by employers, each device measures different dimensions of applicants' qualifications. Whereas competency tests are cognitive ability tests to reveal the intellectual capacity of employees, personality tests aim to uncover personal traits of the workers. Personality tests are generally based on the Five Factor Model, a theory in psychology which classifies human traits into five dimensions. These so called Big Five personality traits are "openness", "conscientiousness", "extraversion", "agreeableness", and "neuroticism". In a laboratory study, Ben-Ner et al. (2004b) relate these five traits to behaviour in dictator games and find that "agreeableness" and "openness" are positively correlated with agents' generosity in gift-exchange games.[6][7] In an earlier contribution Ashton et al. (1998) find that reciprocal altruism as measured in laboratories is closely linked to personality traits "agreeableness" and "emotional stability"; the latter refers to "neuroticism".[8] Day and Silverman (1989) furthermore find that personality traits significantly predict job performance even when cognitive ability is controlled for.

In light of this evidence, I interpret personality tests upon hiring as an indicator for the degree of reciprocity within a firm's labour force – not necessarily because employers explicitly screen for reciprocity but due to correlations between Big Five traits and laboratory measures of reciprocity. In contrast to personality tests, Ben-Ner et al. (2004b) do

[5] This two-step approach of first comparing actual from estimated values and then using these deviations for further inference is similar to Black and Lynch (2001). Black and Lynch, in a first step, estimate production functions for each establishment and then relate the residual to human resource practises.

[6] "Agreeableness" refers to cooperative and compassionate behaviour towards others. "Openness" describes the degree of intellectual curiosity in comparison to cautious behaviour. For further reference, see Atkinson and Hilgard (2000).

[7] A similar finding is reported in Ben-Ner et al. (2004a). Evidence on predictability of Big Five indicators on giving can be found in Ben-Ner and Kramer (2011).

[8] See Part II of this dissertation for an extensive discussion on personality tests and reciprocity. Englmaier and Leider (2012) furthermore provide evidence for increased effectiveness of reciprocal incentives if a personality test beforehand classified agents to be reciprocal.

not find associations between ability and Big Five traits. This suggests that competency tests seem to measure traits which are orthogonal to personality traits like reciprocity.

This study at hand was inspired by a laboratory experiment by Bartling et al. (2012a). In this paper a principal sets a fixed wage and decides on the level of worker's discretion: high discretion implies high productivity of effort at the cost of unlimited shirking opportunities for the employee. Indeed the authors find that providing discretion turns out not to be profitable on its own as shirking is widely prevalent. However combining discretion with high wages renders profitable for employers (and employees) *if* employers can offer such contracts selectively to workers with high effort record. Exogenous variation in the viability of screening agents for their past effort provision allows the authors to identify screening opportunities as the causal determinant of the creation of "good jobs". If screening is permitted then this leads to the emergence of two job-clusters: "good jobs" with high wages, high discretion, and high rent-sharing and "bad jobs" with low wages, low discretion, and little rent-sharing for employees with poor reputation. Most notable, only trust and trustworthiness are necessary for the dichotomy of job-clusters to emerge endogenously.[9]

The most important difference between Bartling et al. (2012a) and this study is that I use *real-world* data. The nature of the data also implies that I cannot exogenously change the market environment as generally done by laboratory studies. Despite this limitations my results give rise to reciprocity as a plausible underlying mechanism casting doubts on efficiency wages and "work ethic" to be the sole explanations of my findings. However, this paper refrains from making causal statements: The study exploits variations which exclusively come from endogenous firms' choices of whether to make use of screening devices or not. Hence this contribution should be regarded as providing correlations between complementarities in HRM practises and labour productivity.

An asset of the WERS 2004 is that the dataset is a representative sample of the British economy, which allows making general statements. This complements the limited generalisability of laboratory experiments, in which the researcher usually does not observe behaviour of real-world decision makers. Of course, real data come at the cost that behavioural responses cannot be measured with the same accuracy in the field as in laboratory experiments: For instance, instead of using effort levels of each employee I rely on firm-wide labour productivity, which itself is a function of employees' effort provision. But regardless whether the relationship between effort and productivity is one-to-one or not, it seems plausible to regard productivity and not merely effort provision as a decisive

[9] Altmann et al. (2013) suggest a non-behavioural mechanism for labour market segregation which is the result of incomplete contracts.

part in managers' objective functions. Despite these potential drawbacks of real-world datasets, my results are remarkably consistent with findings of Bartling et al. (2012a).

Finally this contribution highlights the importance of personality tests in understanding interactions between the behaviour of the employer and the employee in the workplace. Altogether, my results give rise to reciprocity as a plausible explanation for the strong relationship between high firm performance, "good jobs", and screening for personality.

This paper contributes to the rich literature on workplace organisation and personnel economics. Ichniowski and Shaw (2003) provide evidence from various studies that high-performance work systems, in particular high levels of discretion, seems to turn out to be (highly) predictive for firm success whereas traditional and hierarchical workplace organisations are less correlated with favourable outcomes; in this respect this contribution is not an exception. However Ichniowski and Shaw (2003) suggest conventional methods like indoctrination of employees, management culture, or work practises which establish high effort norms and fosters peer pressure to mitigate the free-rider problem. In a study on steel finishing lines, Ichniowski et al. (1995) and Ichniowski et al. (1999) demonstrate that innovative HRM practises like incentive pay and workflows being organised in teams are associated with high line productivity.

In contrast to the above, some recent studies put more emphasis on behavioural aspects of employer-employee relationships. In a laboratory experiment Falk and Fischbacher (2006) provide evidence for workers' adverse behavioural responses after an increase in workplace control, as such policies are perceived as signals for distrust against employees ("hidden costs of control"). Even though such costs may amplify positive effects of high-performance work systems with high discretion (and are consistent with my results), this paper does not build on that mechanism. Another example for employees' non-standard reaction on institutional changes within the workplace is put forth by Nagin et al. (2002) in a field experiment on shirking in call centres. A significant fraction of employees does not react adversely to reductions in monitoring rates suggesting heterogeneity in fairness concerns among workers. These results may imply that choosing the "right" job candidates allows employers to provide discretion to their employees without facing the immediate threat of extensive shirking. This complements findings by Bartling et al. (2012b) who provide an example in which employers with fairness concerns have a broader range of contracts they are able implement as compared to selfish employers.

Similar to Bartling et al. (2012a), this study advances reciprocal responses of employees on employers' behaviour as a potential mechanism to overcome the trade-off between

discretion and shirking.[10] In a theoretical contribution, Englmaier and Leider (2012) incorporate such reciprocal traits of employees into the utility function and solve an otherwise classical moral hazard problem.[11] In their framework, employing reciprocal agents gives employers the opportunity to use two distinct devices to motivate employees: Classical monetary incentives and providing the agent with higher than outside option utility, which induces effort provision through gift-exchange motives.

Closer to this study, however, are empirical papers highlighting the importance of reciprocity in the field. Leuven et al. (2005) provide survey evidence that workers with higher inclination to reciprocity have higher training rates compared to employees with low sensitivity for reciprocity. Similarly, Dohmen et al. (2009) relate individual measures for reciprocity to employee-specific labour market outcomes. They find that positive reciprocity tends to increase wages and is associated with working harder. Bellemare and Shearer (2009) report strong and persistent effects of gift-exchange for a tree-planting firm in British Columbia.[12]

Finally, this paper is related to research on screening methods of employers. Autor and Scarborough (2008) report that screening is widely prevalent among firms. This also is in line with the WERS data used in this study, where one third of all firms use personality tests and even two thirds of surveyed establishments require competency tests upon hiring. In a laboratory experiment, Englmaier et al. (2011) show that employers pay substantial wage premia for information about a worker's cognitive ability and her trustworthiness, implying willingness to pay for screening devices. Finally, Wilk and Cappelli (2003) report that employers differ to a substantial extent in the level they make use of screening devices.

The remainder of the paper is organised as follows: In Section 2 I develop three hypotheses on labour productivity and workplace organisation. Information about the WERS 2004, the matching procedure and the estimation of income and discretion are provided in Section 3, along with an extensive description of the final dataset. Section 4 empirically tests the hypotheses with various econometric specifications. A variety of different specifications to evaluate the robustness of the results is offered in Section 5. Section 6 discusses the results.

[10] Evidence on a relationship between effort and wage can be found in Fehr et al. (1993), Charness (2004), Cohn et al. (2012), and Kube et al. (2012).

[11] Another formal foundation of reciprocity can be found in Falk and Fischbacher (2006).

[12] Kube et al. (2013) find that employees react with negative reciprocal behaviour on wage cuts. Contrary to these studies with positive effects of reciprocity on labour market outcomes, Gneezy and List (2006) do not find strong and long lasting effects of gift-exchange in the field.

2. Hypotheses

In this section I develop testable hypotheses on the relationship between HRM policies and labour productivity. First I generate job-clusters which are derived from combinations of different HRM practises and correspond to at least three major HRM policies: paying efficiency wages, screening for "work ethic" and making use of reciprocity. This classification will later on be refined to provide conditions under which reciprocity appears to be a plausible explanation.

As reviewed in Ichniowski and Shaw (2003), empirical evidence suggests that decentralisation of information flows and implementation of high-performance work systems are associated with firms being more productive. Theoretically, however, the relationship is ex-ante ambiguous as gains from decentralisation may not necessarily outweigh the increase in shirking opportunities for employees which may potentially result in overall lower effort provision. Additional empirical challenges due to the cross-sectional nature of the data arise because firms with high-performance work systems may differ in several observable and unobservable dimensions from firms with rather traditional human resource policies. In this paper I hence regard the relationship between firm performance and the implementation of high-performance work systems as an inherently empirical question which leads me to formulate the first hypothesis rather cautiously.

Hypothesis 1 (Discretion and Firm Performance)**.** *High-performance work systems and firm performance should be positively correlated if productivity gains outweigh potential reductions in effort provision or if firms with high-performance work systems differ in other dimensions which are positively correlated with performance.*

Even if firms with high-performance work systems may report systematically different firm performances, the focus of this paper is to identify HRM-*clusters*, which are correlated with high productivity. Inspired by Bartling et al. (2012a), I focus (next to the degree of employee discretion) on firms' generosity in wage payments and firms' use of screening methods in their hiring practises. For the latter, I distinguish between personality tests which may be correlated with agents' preferences and competency tests which are not.

If neither screening for personality nor the generosity of employees' income is correlated with labour productivity, then the data should not show any pattern in firm productivity between firms with high income provision and/or personality tests and firms without these practises. Such a pattern should be observed, when, firstly, firms are able to motivate employees independently from income levels. For instance, peer effects or

effort norms could generate high effort exertion despite shirking opportunities. If secondly personality tests measure personal traits which are orthogonal to effort provision, then a firm's decision to screen job candidates should not systematically be related to performance.

Following the argument of efficiency wages, it could be sufficient to motivate workers by paying incomes which are substantially higher than obtained in comparable labour relations. This option is disregarded in static principal-agent models, as in these models interactions are one-shot. Hence this implies an abstraction from the threat of dismissing agents if their effort provision is too low. However, efficiancy wages are independent from screening devices as effort provision is not a result of agent's other-regarding preferences but is derived from inter-temporal utility maximisation.

Huang and Cappelli (2010) suggest "work ethic" to be the decisive human resource parameter to attain high firm performance. Employees with high "work ethic" unconditionally provide high effort, which makes it particularly cheap for employers to mitigate the trade-off between discretion and shirking: The employer has to pick job candidates with high "work ethic" but can provide wages that meet the agent's outside option.[13] It follows that firms that screen for personality ("work ethic") and pay low income should perform weakly better than firms paying high income, as wages are, at least in this model, unrelated to effort provision.

Finally, if employees can be motivated via gift-exchange (reciprocity), two conditions have to be met. First, an employee has to have reciprocal traits. By applying personality tests, firms either screen for reciprocity directly or screen for personal traits, which are correlated with reciprocity. The average worker hence should in these firms be more inclined to reciprocity than in firms without screening for personality. But as reciprocity is not an unconditional concept, reciprocal employees only provide high effort if the firm previously provided a gift to them. This is achieved most easily by paying higher than outside option income. Hence, if reciprocity between employers and employees is the underlying mechanism, the data should show particularly high performance for firms which provide high wages *and* screen for personality.

These considerations result in the second hypothesis:

[13] In fact, Huang and Cappelli (2010) find that firms screening for "work ethic" also pay higher wages. The authors explain this finding with competition among firms for these unconditionally motivated workers which finally results in rent-sharing of the generated surplus. Higher wages are hence interpreted as a consequence of high effort and not as a prerequisite for that.

Hypothesis 2 (Job-Clusters and Firm Performance). *Positive associations between high-performance work systems and firm performance should be (weakly) stronger for job-clusters in which the firm*

a) neither screens nor pays high income if employees can sufficiently be motivated otherwise,

b) pays high income if efficiency wages are sufficient to induce high effort from employees (regardless of applying screening devices),

c) screens for "work ethic" and pays low wages, if the firm can pick employees who unconditionally work hard, and

d) screens job candidates for reciprocity and pays high wages, if reciprocity is the underlying behavioural mechanism which induces employees to work hard.

Ex-ante a potential association between "good jobs", personality tests and high firm outcomes, however, is just as plausible as the explanation that screening itself, not narrowly screening for personality is sufficient to predict that pattern. As screening in general is likely to be an indicator of careful human resource policies, firms using tests upon hiring may be able to successfully provide leeway to employees and simultaneously discipline them.

However, if not screening for personality is the decisive factor for the correlation between "good jobs" and firm performance but the fact that firms use any of various possible screening devices, then the same pattern of "good jobs", screening and firm performance should arise similarly for competency tests as well. Firms using any screening devices should be aware of modern human resource practises but only personality tests are associated with traits which predict reciprocity. This insight gives rise to the third hypothesis:

Hypothesis 3 (Competency Tests and Firm Performance). *If screening itself as opposed to narrowly screening for personality is sufficient to generate a positive association between "good jobs" and firm performance, then firms using competency tests to screen job candidates should likewise report exceptional high productivity.*

3. Data

The data comes from the fifth wave of the "Workplace Employment Relations Survey" (WERS 2004) with fieldwork taken place in 2004. Funded by the British government this study is part of a series with previous waves conducted in 1980, 1984, 1990, and 1998

and intents to "provide a nationally representative account of the state of employment relations and working life inside British workplaces".[14] The data consists of four separate datasets including a management survey, an employee survey as well as a questionnaire for employee representatives and a financial performance questionnaire. The following paper studies both the management and the employee survey.

In total, the management survey consists of 2,295 Britain-based establishments from almost all branches of the economy with a minimum of five employees per firm.[15] Within each firm, a maximum of 25 employees were randomly sampled and requested to participate in the employee survey.[16] Overall, in three fourths of the establishments at least one employee returned a questionnaire – hence for 562 firms only information from the management questionnaire is available. Provided a minimum of one employee questionnaire returned, the overall response rate is slightly above 60 percent.

This paper relates firm outcomes to the following three parameters of personnel policy: The generosity of wages, the amount of discretion employees are granted, and whether the firm screens for the personality of job candidates. Whereas I retrieve information on the first two policies from the employee survey, data on personality tests and the outcome variable, labour productivity, are derived from responses by managers from the management survey. In the remainder of this section I first outline the matching procedure followed by an extensive description of the matched dataset with a special focus on the level of discretion employees enjoy as well as their income. Next, I provide information on the procedure of how wages and discretion were estimated on an individual level. I proceed by presenting details on the aggregation procedure and finally an overview of all variables of interest will be provided.

Matching A unique firm identifier allows me to match the management dataset and the employee survey, leading to a dataset which consists of 1,733 firms with 22,451 workers surveyed. The median number of workers surveyed per firm is 13 and the 25th (75th) percentile gives 8 (18) returned questionnaires per firm. Very low response rates and rates close to 25 are rather rare, as can be seen in Figure 4 in Appendix B. The newly generated dataset is weighted with the standard weight which is provided in the employee dataset, is stratified according to the suggested procedure, and standard errors are clustered on

[14] Source: http://www.wers2004.info/wers2004/wers2004.php. January 28, 2013.

[15] Sectors not covered by WERS 2004: Agriculture, hunting and forestry, fishing, mining and quarrying, private households with employed persons, and extra-territorial bodies.

[16] Questionnaires were distributed to all employees if the respective establishment employed less than 25 workers.

firm level.[17]

Description of Employee-Level Variables In the employee questionnaire workers are asked to state their weekly income before taxes and other deductions by marking one out of 14 income intervals, ranging from 50 pounds a week and below to 871 pounds per week and above. The brackets are not equally spaced and the spacing increases with income.[18] Overall 414 employees refused to indicate their income and 19 multi-coded, so that 22,018 answers on income remain. A histogram of the income distribution is depicted in the left panel of Figure 1. The median income in the sample is 311 - 360 pounds a week (16,121 - 18,720 pounds per year). Notice however that employment relations are not necessarily full-time; about one quarter of surveyed employees work less than 30 hours a week.

To proxy for discretion I use two different measures. Employees are asked to rate whether they feel to have "a lot", "some", "a little" or no ("none") influence over what tasks they do in their job and how they do their work.[19] A histogram of both distributions is provided in the right panel of Figure 1. A very large fraction of surveyed employees feels to have at least some leeway over how they perform their work and which actions they can choose. However, even though only approximately 15 (25) percent state to have little or no control over the tasks they do (how they do work) a substantial difference in discretion may exist between statements of having "a lot" or only "some" leeway.

Both income and the individual level of discretion are likely to depend on several characteristics of the employee, such that absolute levels are likely to only be a poor signal about a firm's generosity – the income of an unskilled routine worker with a generous wage will in most cases still be less than that of a badly paid manager. In order to assess generosity in income and discretion, I use deviations of actual income and discretion from estimated values. However, before presenting results from the regressions I first introduce the set of control variables which are summarised in Table 6 in Appendix A.

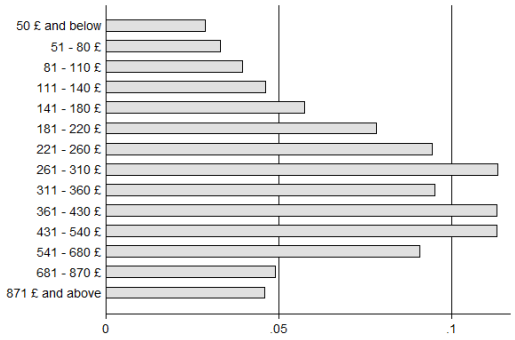
To generate indicators of generous wages and high levels of discretion, I control for an employee's age and tenure, gender, academic and vocational qualifications, whether the employee is a member of a union, as well as the occupational group. Finally, I also control for the number of hours worked per week.

[17] For reference, see http://www.wers2004.info/FAQ.php#5, section 5.6 "How do I apply weights and correctly estimate variances in Stata?", January 28, 2013.

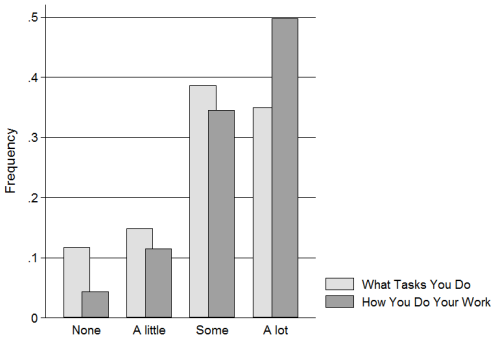
[18] The precise categories of weekly income measured in pounds are as follows: 50 and below, 51 - 80, 81 - 110, 111 - 140, 141 - 180, 181 - 220, 221 - 260, 261 - 310, 311 - 360, 361 - 430, 431 - 540, 541 - 680, 681 - 870 and 871 and above.

[19] The exact wording is "In general, how much influence do you have over the following?", followed by "What tasks you do in your job" ("What Tasks") or "How you do your work" ("How to Work").

Figure 1: Histograms of Income Distribution and Distribution over Perceived Discretion
This figure provides historgrams for the income distribution (measured in pounds per week) (left) and the distribution of two measures for perceived discretion (right) – the dark grey bars of panel (b) refer to preceived discretion in how employees can do their work, the light-coloured bars provide evidence on the autonomy which tasks can be done by employees. All employees who refused to answer or who multi-coded are excluded.



(a) Weekly Income



(b) Discretion

The median age of surveyed employees is between 40 and 49 years, with only 3.3 percent of very young (under 20) and 4.7 percent of old employees (60 and older). Employees are on average between 2 and 5 years in the respective workplace. The data furthermore distinguishes between nine occupational groups. Each employee may belong to: (1) manager and senior officials, (2) professional occupations, (3) associate and technical occupations, (4) administrative and secretarial occupations, (5) skilled trades occupations, (6) caring, leisure and other personal service occupations, (7) sales and customer service occupations, (8) process, plant and machine operatives and drivers, or (9) routine unskilled occupations. As can be seen from Table 6, Appendix A.1, employees are relatively equally sampled from all occupational groups, with slightly less observations for groups (5) - (8). More than a third of the surveyed labour force is unionised and about 17 percent have been in the past. Almost half of the employees (46 percent) obtained a GCSE grade D-G (or comparable)[20], whereas only 4 percent of surveyed employees finished education with a university degree. 80 percent of employees obtained at least a level 1 NVQ (National Vocational Qualification) but less than 0.5 percent reach the top level 5, which involves substantial autonomy and includes bearing high amounts of responsibility. Gender is split almost equally with a slight concentration of males and the average working time is 36 hours per week.

[20] The General Certificate of Secondary Education (GCSE) is a necessary prerequisite for attending high school in the UK education system.

Estimation The estimation of income first has to account for censoring at the bottom and at the top and second has to take bracketing of the answers on income levels into account. For this reason I estimate a variant of a Tobit model which accounts for the survey structure of the data and allows for differently spaced intervals.[21] For the underlying (unrestricted) model of income I assume the following Mincer-type model:

$$w_i = \mathbf{S_i'}\beta_{\mathbf{S}} + \mathbf{F_i'}\beta_{\mathbf{F}} + \mathbf{X_i'}\beta_{\mathbf{X}} + \varepsilon_i. \tag{1}$$

where w_i is the wage employee i receives, $\mathbf{S_i}$ is a vector which contains the set of dummy variables from both academic qualification and vocational training. The vector $\mathbf{F_i}$ includes variables that describe the employee's experience, namely the tenure and the age. $\mathbf{X_i}$ finally summarises all remaining control variables I introduced beforehand, and ε_i is the error term. Tobit models imply two critical assumptions: First, it is assumed that the error is normally distributed and second a homoskedastic error structure is required, i.e. $\varepsilon_i \sim N(0, \sigma^2)$. The first issue is addressed in the robustness section by using log-income instead of income.[22] I deal with the second issue by re-estimating the wage equation with plain OLS, which does not change results.[23]

Results of the wage regression are presented in Table 7, column (1) in Appendix A. Here I provide regression results for a specification that slightly deviates from the classical Mincer wage regression, by omitting the quadratic term on experience. However, both proxies for experience, age of the employee and tenure, are included as dummies for each possible category. As can be seen immediately, an inversely u-shaped relationship between age and income emerges: Ceteris paribus, employees aged between 40 and 49 have the highest income, significantly higher than the base category of under 16 - 17 year olds. With age above 49, the coefficients decrease but income stays significantly above those of job entrants. A similar picture emerges for tenure: Employees with long histories in a particular establishment have substantially higher income than employees with short tenure.

"Occupational group" has the expected influence on income: the more abstract and skill-intensive the occupation, the higher the income. The base category is managers and senior officials. Women on average earn less.

The right panel of Table 7 provides estimates on expected discretion, which I obtain estimating the same equation 1 substituting income with discretion but using simple regression techniques. Column (2) gives results for the question of perceived discretion

[21] For further references see commands "intreg" and "svy" in StataCorp. (2011).
[22] Transforming income into log-income and re-estimation of the model does not change results.
[23] Regression results are available from the author upon request.

concerning what tasks employees are allowed to perform and column (3) refers to how work is done. The pattern here is again similar to what is expected: Older workers with more tenure in more abstract jobs on average report to have more leeway on what tasks to perform and also how to do them.

Aggregation and Generation of Variables Results from estimating equation 1 provides expected income for each employee, conditional on observables. In order to obtain an estimate of whether the employer is generous towards the respective employee with regard to income, I calculate deviations of estimated income from actual income for each employee. As however actual income is only reported within intervals, I use mean income within each interval as actual income. If an employee reports to have income in the highest or lowest category (i.e. her income is censored) I cannot calculate deviations from estimated income, because there is no sensible average for these two categories. If I, for example, set actual income of the top (open) category to its lower boundary, i.e. to 871 pounds per week, then every employee in this category with estimated income higher than 871 pounds would automatically be classified as not being paid generously, because estimated income then always exceeds actual income. In the same manner, all employees falling in the lowest category (and having estimates lower than 50 pounds) would be treated as earning higher than expected income when using the upper end of the category (i.e. 50 pounds per week). For this reason I set observations on income to missing if an employee states to be paid in the highest income category *and* the estimated income is above 871 pounds, which indicates the top (open) category; I proceed analogously with the lowest income category. For discretion I also compare actual and estimated values, but do not have to take censoring into account.

As survey responses for income and discretion are measured in brackets, this procedure is not innocuous. It implies that employees whose continuous estimate for income exactly meets the indicated interval, but exceeds the mean income of the respective interval are classified to receive higher than expected income. But as observed data do not contain any information about within category distribution, these deviations within the interval only provide tendencies towards more or less generosity in salery.

For this reason I alternatively generate a more conservative measure of generosity: Employee contracts are classified to be as expected whenever the estimated value lies in the reported interval. This implies that only large deviations of estimated and actual income lead to contracts being classified as generous. I provide further details on this procedure in Section 5, where I show that results of this paper do not hinge on any of the suggested specifications.

After obtaining measures for firm generosity in income and discretion for each employee, I need to aggregate this information on firm level. This is done by generating the average deviation of employees' income and discretion within each firm, such that each firm obtains a continuous score of generosity separately for income and discretion.

Subsequently I collapse the dataset by deleting all duplicate observations with regard to firms, implying that only firm level information can furthermore be used. All individual employee level information has to be aggregated on firm level at this stage. In order to additionally obtain a more compact measure than continuous firm generosity, I generate a binary variable indicating whether a firm pays higher than expected income/discretion to its employees or not. For that, I calculate the mean deviation across all firms and then relate the score of the respective firm to the average score across all firms.

Description of Firm-Level Variables Summary statistics of firm-level variables, including deviations in estimated averages of employees' income and discretion within a firm are provided in Table 1.

The first panel summarises the distribution of labour productivity, according to self-reported assessments of managers rating their own workplace on a five-point ordinal scale compared to competitors in the same industry.[24] As can immediately be seen, most managers regard the labour productivity of their establishment to be average or even better compared to their industry, whereas only 6.5 percent claim to perform worse. In accordance with recent literature however, this overrating bias from self-reporting seems to merely affect the absolute level but keeps the relative ordering among firms unaffected, as shown in Wall et al. (2004).[25] As the study at hand compares relative productivity and is agnostic about absolute levels, using self-reported productivity measures seems to be valid.

Human resource policy variables are summarised in the second panel of Table 1. The continuous measure of income generosity is centred around zero with 50 percent of the values lying between -49.5 pounds per week (i.e. less than expected) and 39.1. Reducing information to generate a binary variable of deviations in income gives a dummy variable with almost equal split.

[24] The exact question is: "Compared to other establishments in the same industry how would you assess your workplace's labour productivity?" Managers could answer the following: "A lot better than average", "better than average", "about average for the industry", "below the average", "a lot below the average" or "no comparison possible". For intuitive reasons, I re-labelled the variable, such that higher values correspond to higher productivity.

[25] Guthrie (2001) and Baer and Frese (2003) compare subjective and objective performance measures and find product-moment correlations between 0.41 and 0.81. For further evidence on self-report bias see Machin and Stewart (1996).

Table 1: Summary Statistics of Firm-Level Variables

	Obs.	Avg.	SD	Pctl. 25	Pctl. 50	Pctl. 75	Min.	Max.
				Outcome Variable				
Labour Productivity								
A lot better	1977	0.07	0.25	0	0	0	0	1
Better	1977	0.42	0.49	0	0	1	0	1
Average	1977	0.45	0.50	0	0	1	0	1
Worse	1977	0.06	0.24	0	0	0	0	1
A lot Worse	1977	0.004	0.06	0	0	0	0	1
				Modern Human Resource Policy				
Continuous								
Income	1728	−0.89	80.8	−49.5	−8.9	39.1	−329.8	384.7
Income: Binary	1728	0.44	0.49	0	0	1	0	1
Discretion								
What Task	1732	0.01	0.41	−0.21	0.02	0.26	−2.50	1.34
What Task: Binary	1732	0.52	0.50	0	1	1	0	1
How Work	1732	0.01	0.35	−0.18	0.04	0.22	−2.56	1.01
How Work: Binary	1732	0.54	0.50	0	1	1	0	1
Pers. Test	2292	0.34	0.47	0	0	1	0	1
Comp. Test	2291	0.61	0.49	0	1	1	0	1
				Control Variables				
Firm Size	2285	411	947.7	21	67	300	5	10006
Union	2295	0.58	0.49	0	1	1	0	1
Public Sec.	2295	0.27	0.44	0	0	1	0	1
Foreign	2295	0.02	0.14	0	0	0	0	1

Notes: This table provides information on the number of observations, mean and standard deviation, 25th, 50th and 75th percentiles as well as minimum and maximum values of firm performance, human resource practises and control variables. Statistics for each variable are calculated omitting answers "refusal", "don't know" and "not applicable", indicating unclear answers.

Both variables for discretion, i.e. the level of leeway what tasks employees do and how they do their work are centred around zero. As the distributions of both measures for discretion are highly symmetric, the binary representations of discretion have a mean close to 0.5.

About one third of the firms screen job candidates for personality, or, may (indirectly) search for reciprocal types. In contrast, over 60 percent of firms make use of competency tests aiming to elicit workers' (cognitive) ability. The correlation between both screening devices is modest: $\rho = 0.19$. About one third of the firms do not screen at all and one third only uses competency tests. Slightly below 7 percent of the firms only screen for personality and 27 percent rely on both tests. Whereas personality tests are most likely for applicants in high-skilled and abstract jobs (and sales occupations), competency tests

are demanded throughout all occupational groups (see Part II of this dissertation).

The third panel refers to control variables, which aim to control for industry specific differences. More specificly, I control for firm size and firm size squared, unionisation, whether the establishment belongs to the public sector and whether it is foreignly controlled. Furthermore I include dummies for the industry and regional dummies (not part of Table 1).[26]

4. Empirical Analysis

In this section I test for the hypotheses developed in Section 2 with a special focus on Hypothesis 2, by offering three methodological approaches: First I generate HRM-clusters and relate them to firm outcomes. Second, the relationship between outcomes and HRM practises is explored in a fully-fledged regression model. Finally, this paper makes use of continuous deviations of estimated and actual values in income and discretion.

4.1. Discretion and Labour Productivity

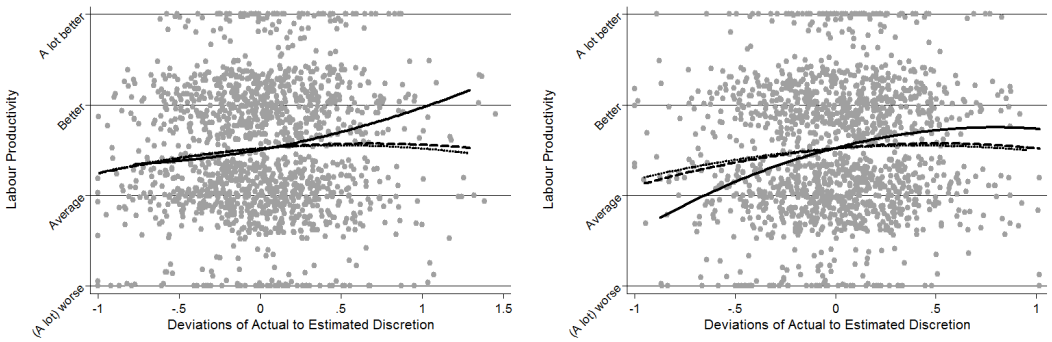
Figure 2 provides an illustrative description of the subsequent empirical elaborations. Raw correlations provide tentative evidence for a positive association between providing high discretion to employees and labour productivity confirming previous results on positive effects of high-performance work systems on firm performance. The dashed line in Panel (a) (discretion measured via the question on leeway of "What Tasks" to perform) refers to the full dataset, i.e. including all firms and shows a slightly increasing but concave pattern. Even though I do not count these graphs as hard evidence, the correlations at least do not support a hypothesis of adverse effects of discretion on labour performance. The second measure for discretion (the level of autonomy of how to perform tasks) shows a similar pattern. A potential interpretation to these findings is that firms do not seem to excessively suffer under shirking, even in cases where employees are granted high discretion and discretion does not come along with specific human resource practises.

Restricting the sample to firms which pay higher than expected income and screen for personality (solid line) draws a different picture. If discretion is at most as high as

[26] Industry is classified according to the UK National Statistics and distinguishes between (1) manufacturing, (2) electricity, gas, and water (3) construction, (4) wholesale and retail, (5) hotel and restaurants, (6) transport and communication, (7) financial services, (8) other business services (9) public administration, (10) education, (11) health, and (12) other community services. Regional dummies are the following: (1) North East, (2) North West, (3) Yorkshire & The Humber, (4) East Midlands, (5) West Midlands (6) East of London, (7) London, (8) South East, (9) South West, (10) Scotland and (11) Wales.

Figure 2: Relationship between Discretion and Firm Performance

These figures illustrate the quadratic relationship between deviations of actual to estimated levels of discretion and labour productivity. The dashed line is based on the entire sample of firms; the solid line describes the relationship for the subset of firms which pay higher than expected income and screen for personality. The dotted line uses the full sample of firms excluding firms which pay high wages and use personality tests (i.e. the subset of firms used for the solid line). For graphical reasons, I use Stata's jitter option for the scatter plot, which adds random noise to observations (the slope of the functions are unaffected).



(a) Discretion: "What Tasks"

(b) Discretion: "How to Work"

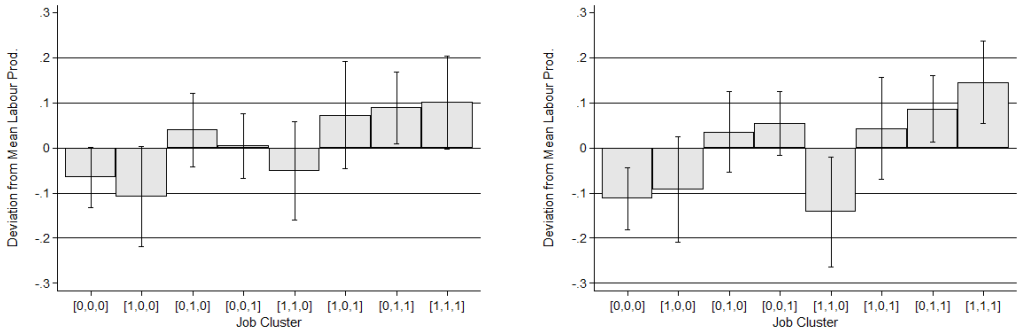
expected, then firms paying high salery and screen job candidates report similar labour productivity as the average firm (panel (a)). But whereas higher discretion only has minor effects on productivity for the average firm, establishments with high saleries and screening devices report higher labour productivity. A different pattern arises when using autonomy on how to perform tasks as measure for discretion, which is depicted in panel (b). The relationship of productivity and discretion of firms which pay high income and screen job candidates is similarly concave but has a steeper slope than the average firm. This implies that if employees in high-income and screening establishments do not receive high levels of discretion, labour productivity sharply decreases. These establishments are successful in terms of labour productivity only if they grant high discretion to their employees.

The dotted line corresponds to the full sample excluding firms which pay high income and screen for personality. I provide these estimates to show that the slight positive slope of the full sample (dashed line) is not driven by the subset of firms, which pay high income and screen for personality (solid line).

Result 1 (Discretion and Firm Performance)**.** *Firms have slightly higher labour productivity if their employees enjoy high levels of discretion.*

Figure 3: Firm Performance and Human Resource Cluster

Panel (a) and (b) provide deviations and 90 percent confidence bands of mean firm performance for firms classified in 8 different HRM-clusters. Read classification as follow: [personality tests, high income, high discretion] i.e. the first position displays whether firms in this cluster use personality tests upon hiring (1) or not (0). The second position refers to whether firms in this cluster pay higher than average expected income (1) or not (0) and the last entry is related to firms with higher (1) or lower (0) than mean expected discretion.



(a) Discretion: "What Tasks"

(b) Discretion: "How to Work"

4.2. Interacting Income, Discretion, and Personality Tests

HRM-Clusters – Graphical Approach In this section I use the binary representation for income and discretion, as explained in Section 3, where each establishment is classified to either pay high or low income and to grant much or little discretion, respectively. I combine this information with firms' screening methods for personality (likewise binary) to assign every single firm to one of eight HRM-clusters. These clusters are all potential combinations of three binary variables implying $2 \times 2 \times 2$ combinations, ranging from firms with (on average) lower than expected wages, lower than expected discretion, and no screening for personality to firms with high wages, high levels of discretion, and personality tests for job candidates.

Figure 3 provides deviations from mean firm performance for each HRM-cluster to the overall performance average of all firms in the dataset.[27] Cluster on the x-axis refer to the following notation: [personality tests, high income, high discretion], where "1" in the respective position implies that firms requires personality tests upon hiring, pay high income or allow for discretion. To the extremes, the illustration shows firms which do not screen for personality, pay low income and do not grant high discretion on the very left of each panel ([0,0,0]) and firms with personality tests, high income and discretion

[27] Relative frequencies of each of the eight HRM-clusters are provided in Figure 5 in the Appendix, Section B.

to the right end ([1,1,1]).

As can immediately be seen in both panels, firms in cluster [1,1,1], i.e. establishments providing "good jobs" and screen for personality, report the highest labour productivity which lies significantly above average productivity. In accordance to Result 1, clusters of both panels are split into two groups: All clusters, in which firms provide higher than estimated discretion perform above average (despite not necessarily significantly higher), whereas firms in 3 out of 4 clusters in both panels with little discretion report to have poorer than average labour productivity. Firms using personality tests and pay high income but do not allow their employees high levels of discretion (cluster [1,1,0]) report exceptionally low labour productivity in panel (b).

HRM-Cluster – Regression Approach Equation 2 provides more structure than simple mean comparisons. I estimate for each firm j the influence of belonging to a certain HRM-cluster, $\mathbf{G_j}$, on ordered outcome variable y_j. The vector $\mathbf{G_j}$ includes seven dummy variables for all but one potential HRM-cluster. $\mathbf{X_j}$ contains control variables, including size of the establishment (number of employees), size squared and dummies for whether the firm is unionised, is owned by a foreign company and belongs to the public sector. Finally, the vector includes dummies for industry and region as described in footnote 26. More precisely, I estimate the following reduced form model:

$$y_j = \mathbf{G_j'}\beta_{\mathbf{G}} + \mathbf{X_j'}\beta_{\mathbf{X}} + \varepsilon_j. \tag{2}$$

Simple regression analysis of equation 2 confirms results from previous mean comparisons.[28] Using firms without personality tests, low income and limited discretion, i.e. cluster [0,0,0] as base category (columns (1) and (3)), the job-cluster which refers to "good jobs" and screening ([1,1,1]) predicts significantly higher labour productivity for both measures of discretion. In accordance to panel (b) in Figure 3, all three remaining job-clusters with high discretion concerning employees' autonomy on how to carry out their work ([0,0,1], [1,0,1] and [0,1,1]) yield significantly higher labour productivity as compared to cluster [0,0,0] (column (3)).

[28] Throughout this paper I use simple regression techniques, because an ordered probit approach yields qualitatively the same results. As simple regression analysis facilitates the interpretation, I decided to report these results. Ordered probit results are available from the author on request.

Table 2: Regression of Labour Productivity on HRM-Cluster

	What Tasks		How Work	
	(1)	(2)	(3)	(4)
HRM Cluster				
[0,0,0]		−0.50***		−0.50***
		(0.14)		(0.16)
[1,0,0]	0.100	−0.40**	0.13	−0.36*
	(0.15)	(0.18)	(0.15)	(0.20)
[0,1,0]	0.21**	−0.29**	0.29***	−0.20
	(0.10)	(0.14)	(0.10)	(0.17)
[0,0,1]	0.13	−0.37***	0.26***	−0.24
	(0.10)	(0.14)	(0.093)	(0.16)
[1,1,0]	−0.065	−0.57**	0.019	−0.48*
	(0.25)	(0.26)	(0.25)	(0.28)
[1,0,1]	0.21	−0.29*	0.32**	−0.17
	(0.14)	(0.17)	(0.13)	(0.19)
[0,1,1]	0.13	−0.37***	0.16*	−0.33**
	(0.11)	(0.14)	(0.098)	(0.16)
[1,1,1]	0.50***		0.50***	
	(0.14)		(0.16)	
Union	−0.064	−0.064	−0.042	−0.042
	(0.084)	(0.084)	(0.084)	(0.084)
Pub. Sector	−0.068	−0.068	−0.060	−0.060
	(0.11)	(0.11)	(0.11)	(0.11)
Foreign	0.55***	0.55***	0.64***	0.64***
	(0.12)	(0.12)	(0.13)	(0.13)
Constant	3.61***	4.11***	3.52***	4.02***
	(0.20)	(0.22)	(0.20)	(0.24)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1812	1812	1812	1812
R^2	0.093	0.093	0.103	0.103

Notes: This table provides linear regression coefficients and standard errors of labour productivity on HRM cluster and controls. The first panel (column (1) and (2)) refers to answers on the question "What Tasks" as proxy for discretion, columns (3) and (4) use "How to Work". Columns (1) and (3) use cluster [0,0,0] as base category; column (2) and (4) omit cluster [1,1,1].
Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Columns (2) and (4) provide results for the same regression but relative to cluster [1,1,1] as baseline category. It is important to notice that all job-clusters yield significantly worse labour productivity, when discretion refers to which tasks employees are allowed to do (column (2)). This implies that also within the subset of rather successful firms using high discretion (Result 1), paying high wages and *simultaneously* screening for personality is

associated with significantly higher labour productivity.

This result similarly applies to the second measure of discretion. Although a few job-clusters are not significantly distinguishable from job-clusters of "good jobs" and personality tests, point estimates are exclusively negative.

Fully-Fledged Interaction Model In the following, I depart from the rather inflexible analysis of eight pre-specified job-clusters and allow for a fully-fledged interaction model of personality tests, income and discretion. To facilitate interpretation the following analysis is based on binary HRM measures. Subsequently I allow HRM practises to be measured continuously.

The following model measures the relationship of high income ($I_j = 1$), high discretion ($D_j = 1$), and compulsory personality tests upon hiring ($P_j = 1$) on firm performance y_j for each firm j. Apart from the main effects of HRM practises on performance, interaction terms are of main interest as these effects provide insight into complementarities of HRM practises: $I_j \times D_j$ is defined as the interaction between income and discretion and yields value 1 if firm j both provides high income and leaves discretion to their employees (provide "good jobs"). $I_j \times P_j$, the interaction between high income and compulsory personality tests and $D_j \times P_j$, which interacts high discretion with personality tests are defined accordingly. The three-way interaction $I_j \times D_j \times P_j$ is equal to 1 for all firms which were previously classified to provide "good jobs" *and* additionally demand personality tests when recruiting new employees. Finally, the vector $\mathbf{X_j}$ contains control variables as defined previously in this section.

$$y_j = \beta + \beta_I I_j + \beta_D D_j + \beta_P P_j + \beta_{ID}(I_j \times D_j) + \beta_{IP}(I_j \times P_j) + \beta_{DP}(D_j \times P_j) + \beta_{IDP}(I_j \times D_j \times P_j) + \mathbf{X_j'}\beta_{\mathbf{X}} + \varepsilon_j \quad (3)$$

Table 3 provides estimation results of model 3 using simple regression analysis.[29] As all HRM practises are defined on a binary support, the constant can intuitively be interpreted as the labour productivity of the average firm (average with respect to controls) which neither pays high income nor provides discretion nor screens for personality. The main effect of income is the difference in reported labour productivity of firms which do not pay high income compared to establishments which do pay high income (everything else equal). If the ceteris paribus condition implies that firms neither grant high levels of discretion nor screen for personality then the coefficient on income displays the full effect of high income on productivity. The positive and significant coefficient on income for

[29] Here, again, applying ordered probit estimation qualitatively does not change the results.

Table 3: Regressions of Labour Productivity on HRM Complementarities

	What Tasks		How Work	
	(1)	(2)	(3)	(4)
High Income (I)	0.24**	0.21**	0.29***	0.29***
	(0.11)	(0.10)	(0.11)	(0.10)
High Discretion (D)	0.18*	0.13	0.31***	0.26***
	(0.10)	(0.10)	(0.099)	(0.093)
Pers. Test (P)	0.12	0.100	0.15	0.13
	(0.15)	(0.15)	(0.15)	(0.15)
I × D	−0.24	−0.22	−0.37**	−0.39***
	(0.15)	(0.15)	(0.15)	(0.15)
I × P	−0.37	−0.37	−0.40	−0.41
	(0.29)	(0.29)	(0.30)	(0.29)
D × P	−0.041	−0.024	−0.089	−0.066
	(0.21)	(0.20)	(0.21)	(0.20)
I × D × P	0.71**	0.67*	0.72*	0.67*
	(0.36)	(0.35)	(0.38)	(0.36)
Union		−0.064		−0.042
		(0.084)		(0.084)
Pub. Sector		−0.068		−0.060
		(0.11)		(0.11)
Foreign		0.55***		0.64***
		(0.12)		(0.13)
Constant	3.39***	3.61***	3.34***	3.52***
	(0.073)	(0.20)	(0.069)	(0.20)
Firm Controls	*No*	*Yes*	*No*	*Yes*
Subpop. Observations	1815	1812	1815	1812
R^2	0.023	0.093	0.037	0.103

Notes: This table provides linear regression coefficients and standard errors of labour productivity on binary variables of income, discretion, personality tests and its interactions. The first panel (column (1) and (2)) refers to answers on the question "What Tasks" as proxy for discretion, columns (3) and (4) use "How to Work".
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

all specifications hence provides evidence that firms paying high income ceteris paribus report higher levels of labour productivity as compared to firms which pay low income. Similarly a positive and significant coefficient on discretion in 3 out of 4 specifications reflects a positive correlation between discretion and labour productivity. Only the main effect of personality tests is not significantly different from zero.[30]

[30] In Part II of this dissertation, Englmaier, Kolaska, and Leider in fact find a positive relationship between firm performance and personality tests. As model specifications are very different, results are not directly comparable.

The two-way interactions $I_j \times D_j$ is significantly negative for discretion measured in employees' leeway how to work. This means that firms paying high income and leave their workers discretion but do not screen for personality report significantly lower productivity of labour than firms with limited discretion all other parameters equal. Applying a causal interpretation, $I_j \times D_j$ is the additional effect from complementarities between income and discretion when switching from low to high wages in firms which already provide high levels of discretion. As the interaction is significantly negative, this policy, which corresponds to paying efficiency wages, turns out to be not successful. The interaction of high income and personality tests (paying low income) $D_j \times P_j$ is not significantly distinguishable from zero, discarding optimality of the strategy to rely solely on "work ethic" but refrain from paying high wages. Finally the third interaction, $I_j \times P_j$ does similarly not depict a significant pattern for any measure of discretion.

The main interest however, is placed on the three-way interaction term between income, discretion and personality tests, $I_j \times D_j \times P_j$. This interaction is significant in all four specifications. Intuitively (and causally interpreted) this interaction is the additional effect of introducing personality tests given that the respective firm already pays high income and leaves their employees high discretion on productivity. The overall effect of personality tests (given the firm offers "good jobs") is the sum of all coefficients, which contain personality tests, i.e. $\beta_P + \beta_{IP} + \beta_{DP} + \beta_{IDP}$. A Wald test on the null hypothesis that the sum of all four coefficients containing personality tests is zero can be rejected in three out of four specifications.[31]

Analysis with Continuous Deviations A potential immediate critique of the previous analysis is that it entirely relies on the binary classification of firms' HRM practises, which implies that a lot of information is (unnecessarily) lost. In the following, I exploit continuous deviations of actual income and discretion to estimated outcomes. However, this procedure also comes at a cost. As answers on the survey question concerning discretion are ordinal, deviations are not directly interpretable, suggesting a very cautious interpretation of the subsequent results.

In order to address complementarities of different HRM practises, it is essential to include interactions, as done before. However as multiple interaction effects of continuous variables are very problematic to interpret, I keep the analysis tractable by dividing the dataset into firms using personality tests upon hiring ($P_j = 1$) and those who do not ($P_j = 0$). For each dataset $k \in \{0, 1\}$, I then separately estimate the effect of income

[31] The models which include firm controls (columns (2) and (4)) strongly reject the null at a 1%-significance level and at a 5%-significance level, respectively. Specification (1) rejects the null at a 10%-level and model (3) fails to reject it (Prob $>$ F $= 0.18$).

i_j, discretion d_j and the multiplication term $(i_j \times d_j)$ of income and discretion on firm performance y_j.

I estimate the following reduced-form model:

$$y_{j|P_j=k} = \beta_i i_j + \beta_d d_j + \beta_{id}(i_j \times d_j) + \mathbf{X_j'}\beta_{\mathbf{X}} + \varepsilon_j \qquad \forall k \in \{0,1\}. \tag{4}$$

Equation 4 is estimated with simple regression techniques.[32] As both income and discretion are centred around zero, the intercept of each regression depicts mean labour productivity if actual and estimated values of income and discretion are exactly aligned (meaning that firms exactly provide wages and discretion as expected), at the mean of all control variables. β_d can be interpreted as the average difference in reported labour productivity of the mean firm (with regard to controls) if employees on average report one step higher perceived discretion and income is exactly as expected (i.e. $i_j = 0$, leading the interaction term to be zero). The same is true for $i_j > 0$ and $d_j = 0$; the main effect of income can be seen as the impact of paying one pound more per week on average to each employee on labour productivity, given that the firms allow discretion exactly as expected. The interaction term is hence only different from zero, if both HRM measures depart from expected values. A positive interaction can be interpreted as follows: The more actual income exceeds estimated income, the stronger the association of discretion on firm performance becomes. The interpretation vice versa is true as well, implying that the more "excess" (i.e. above estimated) discretion firms allow their employees on average, the more increasing income is associated with firm outcomes. As in a framework with continuous wage and measures for discretion no firm exhibits income and discretion exactly at the sample mean, the effect of income on firm output depends on the level of discretion if the interaction effect is significantly different from zero.

Columns (1) and (3) of Table 4 provide evidence on estimates for income, discretion and the interaction of both for the subset of firms, which do not use personality tests when screening job candidates. For this subset of firms, neither income nor discretion (or the interaction) is associated with labour productivity. This is true for control variables as well with the exception of firms which are foreign owned: These firms report significantly higher labour productivity.

Restricting the dataset to firms which screen for personality reduces the number of observations, as can be seen in column (2) and column (4). Whereas the main coefficients of income increase but are indistinguishable from zero even if firms screen for personality, coefficients of discretion sharply increase 10-fold and 7.5-fold. Likewise point estimates

[32] Here, again, applying ordered probit estimation does not alter the results.

Table 4: Regression of Labour Productivity on HRM Complementarities II

	What Tasks		How Work	
	w/o PT	w/ PT	w/o PT	w/ PT
	(1)	(2)	(3)	(4)
Income	−0.000086	0.00066	−0.00022	0.00070
	(0.00059)	(0.00075)	(0.00057)	(0.00074)
Discretion	0.026	0.30*	0.064	0.48**
	(0.094)	(0.17)	(0.089)	(0.24)
Income × Discretion	−0.00094	0.0019	−0.0011	0.0044**
	(0.00065)	(0.0015)	(0.00073)	(0.0019)
Union	−0.042	−0.13	−0.024	−0.13
	(0.095)	(0.18)	(0.097)	(0.17)
Pub. Sector	−0.095	0.052	−0.093	0.015
	(0.12)	(0.20)	(0.12)	(0.20)
Foreign	0.48***	0.88***	0.50***	0.59**
	(0.14)	(0.24)	(0.14)	(0.24)
Constant	3.75***	3.71***	3.73***	3.70***
	(0.22)	(0.27)	(0.22)	(0.28)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	998	496	998	496
R^2	0.088	0.249	0.092	0.266

Notes: This table provides linear regression coefficients and standard errors of labour productivity on income, discretion and the interaction between income and discretion. The first panel (column (1) and (2)) uses answers on the question "What Tasks" as a proxy for discretion, columns (3) and (4) use "How to Work". Columns (1) and (3) correspond to the restricted set of firms which do not use personality tests, columns (2) and (4) refer to establishments which make use of personality tests for job candidates. Level of Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of the interaction term increase compared to firms without personality tests and result in significant correlation of the interaction between discretion and income, and labour productivity in column (4). Using leeway on what task to perform as measure for discretion and conditioning on firms with personality tests also increases point estimates of the interaction term, as can be seen in column (2). Higher standard errors however, lead to insignificant coefficients, which may be a result of limited sample size. It is important to notice that also the goodness-of-fit is approximately three times as high when conditioning the sample on firms with personality tests but otherwise estimating the identical model. This provides evidence for high explanatory power of HRM practises for labour productivity.

To show that the impact of discretion on labour productivity crucially depends on the

level of income if firms screen for personality, I finally provide estimates for the evolution of discretion along different levels of income deviations. Estimates in Table 5 are simply the linear combination of the main effect of discretion on labour productivity for three different levels of income: income evaluated at its mean (i.e. no deviation from estimated income), actual income being one standard deviation above and one standard deviation below estimation. The first panel provides estimates for the subset of firms, which screen for personality and the second panel for firms not using compulsory personality tests.

The effect of discretion on labour productivity is largest when income is high and drops considerably with low income for firms which screen job candidates for personality. This is true for both measures of discretion. The coefficient for discretion, given that income is one standard deviation below its mean, is not distinguishable from zero; in that case an increase in discretion does not predict higher labour productivity. As before, the subset of firms without personality tests do not show a clear pattern. I interpret this as personality tests being crucial for "good jobs" to translate into high labour productivity.

Result 2 (Job-Clusters and Firm Performance)**.** *Firms paying higher than expected income to their employees while at the same time allowing their workforce substantial discretion on how to perform tasks (HRM practises which are associated with "good jobs") report significantly higher labour productivity. However, this is the case only if these firms screen job candidates for personality. Hence, this pattern is consistent with reciprocity as underlying mechanism to achieve high firm performance, discarding explanation (a) no additional human resource practise necessary, (b) efficiency wages only and (c) relying on "work ethic" as sole explanations.*

4.3. Competency Tests

Firms offering "good jobs" report to have significantly higher labour productivity if they screen for personality when hiring employees. However, personality tests may only be a proxy for advanced HRM practises which itself is likely to be positively correlated with firm performance. Competency tests may similarly proxy for firms with advanced HRM policies but ability tests unlike personality tests aim not to uncover personal traits of the employee but try to reveal her ability. Though both tests are indicators of modern HRM practises of a firm, only personality tests are consistent with reciprocity as an enhancing mechanism for productivity. Hence, this section summarises the results using the same analysis as in the previous section but defining competency tests instead of personality tests as a screening device.

Table 8 (Appendix, Section A.2) provides estimates from equation 2, where personality

Table 5: Impact of Discretion on Labour Productivity

	Coeff.	Std. Err.	t	p>|t|	95% Conf.	Int.
With Personality Test						
What Tasks						
Slope at Mean of Income	0.30	0.17	1.78	0.08	−0.03	0.64
Slope at Mean + 1 SD of Income	0.46	0.23	2.01	0.05	0.01	0.91
Slope at Mean − 1 SD of Income	0.15	0.18	0.80	0.42	−0.21	0.50
How Work						
Slope at Mean of Income	0.48	0.24	1.98	0.05	0.005	0.95
Slope at Mean + 1 SD of Income	0.83	0.37	2.26	0.02	0.11	1.55
Slope at Mean − 1 SD of Income	0.12	0.17	0.74	0.46	−0.20	0.45
Without Personality Test						
What Tasks						
Slope at Mean of Income	0.03	0.09	0.28	0.78	−0.16	0.21
Slope at Mean + 1 SD of Income	−0.05	0.12	−0.41	0.69	−0.29	0.19
Slope at Mean − 1 SD of Income	0.10	0.09	1.12	0.26	−0.08	0.28
How Work						
Slope at Mean of Income	0.06	0.09	0.73	0.47	−0.11	0.24
Slope at Mean + 1 SD of Income	−0.03	0.12	−0.21	0.83	−0.27	0.22
Slope at Mean − 1 SD of Income	0.15	0.09	1.81	0.07	−0.01	0.32

Notes: This table provides linear regression coefficients, standard errors and confidence intervals for the effect of discretion on labour productivity for different levels of income. The first line of each panel is the compound effect (main effect plus interaction) of discretion at mean income payment. The second line describes the effect of discretion on productivity if income is one standard deviation above the mean. The third line corresponds to estimates for income one standard deviation below the mean. The first panel reports estimates for firms using personality tests, the second panel for firms without screening for personality.

tests P_j are replaced by competency tests C_j. Column (1) and (3) report significantly higher labour productivity if firms provide "good jobs" and screen for ability compared to firms in cluster [0,0,0]. Furthermore using firms with "good jobs" and ability tests as base category I find significantly higher levels of labour productivity compared to establishments in other job-clusters. This effect is particularly strong for discretion measured by the question on how autonomous employees are allowed to perform their tasks. Even though results in these regressions are considerably weaker than results on personality tests, using this piece of evidence does not speak against an association of labour productivity and "good jobs" combined with competency tests.

Results on estimating three-way interactions of model 3 are provided in Table 9. Discretion relates positively to labour productivity and the coefficient is substantial if discretion is measured by employees' leeway on how to work. More importantly however,

the interaction effect of high income, high discretion, and competency tests is indistinguishable from zero for all four specifications. Intuitively this means that firms providing "good jobs" and screen for competency when hiring employees do not report higher labour productivity than firms offering equally good jobs but do not screen job candidates for ability. This is in sharp contrast to findings on the impact of personality tests which renders screening itself unlikely to be the driving force of "good jobs" to translate into high labour productivity.

Result 3 (Competency Tests and Firm Performance)**.** *Screening job candidates for competency as opposed to personality has considerably less power to explain labour performance if firms offer "good jobs". Not screening itself but screening for personality is decisive for "good jobs" to turn into high labour productivity.*

5. Robustness

In this section I perform a series of robustness checks: First, instead of estimating deviations in income and discretion, I use raw responses on income and discretion. Secondly, I only classify firms to be generous if actual income/discretion is at least one category higher than estimated values. In a third check, I perform the Tobit model substituting income with logarithmised income. Fourth, in order to provide a different method of how to aggregate employee responses into firm averages, I first calculate binary firm generosity towards each employee and second aggregate over all employees per establishment. To address the fact that some firms simultaneously make use of personality and ability tests, I finally re-estimate models on both screening devices excluding firms which simultaneously search for personality and ability.

Raw Responses In this section, I re-estimate models 2 and 3 using raw responses from the employee questionnaire. By skipping the entire procedure on estimating income and discretion I hence do not account for personal characteristics of each employee. However, because firms differ in the composition of *which* employees answer the questionnaire I may systematically over- and underrate income and discretion of firms. If questionnaires however were returned purely random in each establishment, then using raw correlations should only increase noise but is not expected to systematically bias coefficients.

As shown in Table 10, column (1) and (2) in the Appendix, Section A.3, estimates for both measures of discretion show a similar pattern as in the main section, even though coefficients are estimated less precisely. For model 2 and using job-clusters of firms providing "good jobs" and personality tests as reference category we find consistently negative

coefficients; some clusters exhibit significant lower associations with labour productivity. Column (1) and (2) in Table 12 finally re-estimates model 3 substituting estimated income and discretion with raw values. Here, we find significant interaction terms between "good jobs" and personality tests only for one of two specifications.

Categorising Deviations Another potential critique could be the use of (hypothetical) continuous deviations between estimated and actual values for income and discretion. For the previous analysis I generate a continuous measure of expected income, though actual income is only observed within intervals. This, however, implies that firms are classified to pay less generous income to a certain employee if the estimated income is higher than the average income within the respective interval. Some employees hence will falsely be categorised to receive generous wages. The situation is analogously when estimated income is below the mean income within an interval, which leads firms to appear generous. Despite inaccuracies, however, the procedure so far is not expected to systematically bias the results (assuming a symmetric income distribution around the mean of each category) but having rather a tendency to increase the variance.

To nevertheless address these concerns I subsequently classify firms only then to be generous with regard to income if the lower bound of their actual income interval exceeds their estimated income. All employees, whose estimated income lies in the interval of their actual payments are classified as to receive income as expected. If firms pay as expected then deviations are equalised to be zero; if establishments pay higher (lower) than expected then I use the mean payment of the actual interval as reference payment. I similarly proceed with estimates on discretion, where employees' answers on perceived discretion are scales from "1" ("none") to "4" ("a lot").[33] Given this scale, I only count answers to be higher (lower) than expected if expected and actual values deviate by at least the positive (negative) magnitude of 0.5. In these cases actual discretion is not the closest integer to estimated discretion.

Table 10 (column (3) and (4)) provides estimates from model 2 for both suggested measures for discretion. Contrary to the main results, in column (3) the cluster [1,1,1] is not significantly stronger correlated with productivity compared to some other cluster, though all point estimates direct towards this relationship. Using responses on the question "How to Work" as measure for discretion, estimates are comparable to the main results (column (4)). A similar picture arises for model 3 in Table 12. Human resource indicators do not show significant patterns (column (3)) using "What Tasks" as mea-

[33] Issues on the interpretability of ordinal responses are dealt with in "Alternative Method of Aggregation" in this section.

sure for discretion – the magnitude of the coefficient on "good jobs" in combination with screening for personality is considerable smaller than in the base specification. Applying "How to Work" as measure for discretion, results are again comparable to the main section.

Log-Income Crucial for the estimation of income with interval regressions is the normality assumption. As the distribution of log-income in many applications is closer to the Gaussian normal (compared to raw income) I re-estimate model 1 after transforming income into log-income.

The transformation has no effect on the results of job-clusters (Table 10 in Appendix A.3): Relating labour productivity to HRM-clusters I find negative point estimates for all job-clusters when omitting firms with "good jobs" and personality tests as base category. Six out of seven clusters (for each measure of discretion) report significantly lower productivity. Results on model 3 are presented in Table 12. Here, however, I do not find significant interaction effects on $I_j \times D_j \times P_j$ for discretion measured by the question on "What Tasks". Using "How to Work", I find high and significant correlations between the three-way interaction and labour productivity. Hence, the findings in this section affirm that results from the main section are robust to a transformation of income into log-income.

Alternative Method of Aggregation A serious concern of the previous analysis is the interpretation of deviations from estimated HRM practises, although discretion is an ordinal measure without scale. Moreover, when calculating firm averages, exceptionally high positive (negative) deviations of discretion for one employee could potentially offset lower (higher) than expected levels for several employees.

Here, I address both problems by changing the order of aggregation and averaging across firms. Before, I summed deviations in income (discretion) across all employees within one establishment and then took the mean deviation in income (discretion) within the firm. In this section I first calculate a binary measure for each single employee, indicating whether this respective employee receives generous income (discretion) or not from her firm. In a second step I calculate the fraction of employees with high income (discretion) for each firm. Finally, a firm is classified to be generous with regard to income (discretion) if it provides more employees high income (discretion) than the average firm in the sample.

Estimates qualitatively do not change when applying the latter method of aggregation. Columns (1) and (2) in Table 11 provide estimates for model 2 using job-cluster [1,1,1]

as base category. As seen immediately almost all job-clusters yield significant negative correlations with labour productivity compared to firms which provide "good jobs" and screen for personality. In Table 13 (columns (1) and (2)) I find positive and significant correlations of productivity with the interaction of "good jobs" and screening, $I_j \times D_j \times P_j$. This implies that the effect of HRM practises on productivity is not driven by the way how I aggregate information from the regressions on income and discretion onto firm level.

Excluding Firms with Personality and Competency Tests A final potential worry of the previous analysis is the fact that some firms use personality tests and competency tests simultaneously. In this paragraph I exclude exactly these firms from the analysis. This may be particularly interesting for the results on competency tests because these results include a subset of firms which also screen for personality.

Estimation results on job-clusters excluding the subset of firms using both screening devices are provided in Table 11, where columns (3) and (4) refer to personality tests as a screening device and columns (5) and (6) to ability tests. Comparing results to the main tables, I do not find qualitative differences. This is only partly true for model 3. As can be seen in columns (3) and (4) the interactions of "good jobs" and personality tests are not significantly different from zero for both measures of discretion, even though point estimates sharply increase. This observation however, may be explained by limited sample size resulting in standard errors which are approximately 50 percent higher compared to the base specification. Finally, competency tests (columns (5) and (6)) in combination with "good jobs" cannot explain labour productivity, which is in line with Result 3. Here, standard errors are comparable to the standard errors in the main regressions.

6. Discussion

Research in personnel economics has highlighted the importance of workplace organisation for firm success. Rather recently, however, a number of studies find that behavioural aspects within firms may shape outcomes. This implies that taking the "right" actions may allow employers to benefit from non-standard behaviour of employees. Three of these potentially "right" actions are presented in this paper.

This paper uses field evidence from the "Workplace Employment Relations Survey" to relate three human resource policies – paying high income to employees, leaving worker high discretion, and screening for personality or competency – to firm performance. I

show that firms which pay high income, grant high level of discretion, and screen their job candidates for personal traits report to have exceptional high labour productivity. I interpret this finding as evidence consistent with employees responding to gift-exchange; job-clusters which are associated with efficiency wages or high "work ethic" of employees alone are not associated with positive firm outcomes. This similarly applies for firms which screen for ability instead of personality.

In a broader context, this analysis shows the importance of personality tests when screening job candidates. Interestingly, however, in this dataset only one third of the firms make use of that screening device, which is a bit of a puzzle: If personality tests are the key to increase labour productivity (because this device reduces the number of employees with adverse behaviour towards the firm) then one should expect firms to increasingly make use of screening for personality. If however, only a limited number of employees in the population exhibit reciprocal traits then rising demand for these workers could lead to segmentation in the labour market: Successful firms with reciprocal employees and "good jobs" on one side and firms providing jobs with low payments and low discretion on the other. (The argument of segmentation in labour markets has, in a slightly different context, already been made by Bartling et al. (2012a))

A natural next step could be to provide causal evidence of HRM practises, in particular of personality tests on firm performance using field data. Whereas laboratory studies can isolate underlying principles, it is often not clear whether the identified mechanism has real-world implications. Gaining evidence on the actual importance of reciprocal behaviour between employer and employee could improve labour market relations with potential benefits for both parties.

References

Altmann, S., A. Falk, A. Grunewald, and D. Huffman (2013): "Contractual Incompleteness, Unemployment, and Labor Market Segregation," *mimeo*.

Ashton, M. C., S. V. 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.

Atkinson, R. L. and E. R. Hilgard (2000): *Hilgard's introduction to psychology*, Fort Worth and Tex: Harcourt College Publishers, 13 ed.

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

Baer, M. and M. Frese (2003): "Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance," *Journal of Organizational Behavior*, 24, 45–68.

Bartling, B., E. Fehr, and K. M. Schmidt (2012a): "Screening, Competition, and Job Design: Economic Origins of Good Jobs," *American Economic Review*, 102, 834–864.

——— (2012b): "Use and Abuse of Authority: A Behavioral Foundation of the Employment Relation," *SSRN Electronic Journal*.

Bellemare, C. and B. Shearer (2009): "Gift giving and worker productivity: Evidence from a firm-level experiment," *Games and Economic Behavior*, 67, 233–244.

Ben-Ner, A., F. Kong, and L. Putterman (2004a): "Share and share alike? Gender-pairing, personality, and cognitive ability as determinants of giving," *Journal of Economic Psychology*, 25, 581–589.

Ben-Ner, A. and A. Kramer (2011): "Personality and altruism in the dictator game: Relationship to giving to kin, collaborators, competitors, and neutrals," *Personality and Individual Differences*, 51, 216–221.

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

Black, S. E. and L. M. Lynch (2001): "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity," *Review of Economics and Statistics*, 83, 434–445.

Charness, G. (2004): "Attribution and Reciprocity in an Experimental Labor Market," *Journal of Labor Economics*, 22, 665–688.

Cohn, A., E. Fehr, and L. Goette (2012): "Fair Wages and Effort: Evidence from a Field Experiement," *mimeo*, 2012.

Day, D. and S. Silverman (1989): "Personality and Job Performance: Evidence of Incremental Validity," *Personnel Psychology*, 42, 25–36.

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

Englmaier, F. and S. Leider (2012): "Managerial Payoff and Gift Exchange in the Field," *SSRN Electronic Journal*.

Englmaier, F., S. Strasser, and J. Winter (2011): "Worker Characteristics and Wage Differentials: Evidence from a Gift-Exchange Experiment," *mimeo*.

Falk, A. and U. Fischbacher (2006): "A theory of reciprocity," *Games and Economic Behavior*, 54, 293–315.

Fehr, E. and S. Gächter (2000): "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, 14, 159–182.

Fehr, E., G. Kirchsteiger, and A. Riedl (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation," *The Quarterly Journal of Economics*, 108, 437–459.

Fehr, E. and K. M. Schmidt (2006): "Chapter 8 The Economics of Fairness, Reciprocity and Altruism – Experimental Evidence and New Theories," in *Handbook of the Economics of Giving, Altruism and Reciprocity, Volume I*, ed. by S.-C. Kolm and J. Mercier Ythier, Elsevier, vol. 1, 615–691.

Gneezy, U. and J. A. List (2006): "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments," *Econometrica*, 74, 1365–1384.

Guthrie, J. (2001): "High-Involvement Work Practices, Turnover, and Productivity: Evidence from New Zealand," *Academy of Management Journal*, 44, 180–190.

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

Ichniowski, C. and K. Shaw (2003): "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices," *Journal of Economic Perspectives*, 17, 155–180.

Ichniowski, C., K. Shaw, and R. W. Crandall (1995): "Old Dogs and New Tricks: Determinants of the Adoption of Productivity- Enhancing Work Practices," *Brookings Papers on Economic Activity. Microeconomics*, 1995, 1.

Ichniowski, C., K. Shaw, and G. Prennushi (1999): "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 87, 291–313.

Kube, S., M. A. Maréchal, and C. Puppe (2012): "The Currency of Reciprocity: Gift Exchange in the Workplace," *American Economic Review*, 102, 1644–1662.

——— (2013): "Do Wage Cuts Damage Work Morale? Evidence from a Natural Field Experiment," *Journal of the European Economic Association*, forthcoming.

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

Machin, S. and M. Stewart (1996): "Trade Unions and Financial Performance," *Oxford Economic Papers*, 48, 213–241.

Nagin, D. S., J. B. Rebitzer, S. Sanders, and J. T. Lowell (2002): "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment," *American Economic Review*, 92, 850–873.

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

StataCorp. (2011): "Stata 12 Base Reference Manual," *College Station, TX: Stata Press*.

Wall, T. D., J. Michie, M. Patterson, S. J. Wood, M. Sheehan, C. W. Clegg, and West Michael (2004): "On the Validity of Subjective Measures on Company Performance," *Personnel Psychology*, 57, 95–118.

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

Appendix

A. Tables

A.1. Data Description

Table 6: Summary Statistics of Employee Characteristics

	Obs.	Avg.	SD	Min.	Max.
Age					
< 20	22362	0.03	0.18	0	1
20 - 29	22362	0.18	0.39	0	1
30 - 39	22362	0.25	0.43	0	1
40 - 49	22362	0.27	0.44	0	1
50 - 59	22362	0.22	0.41	0	1
60+	22362	0.05	0.21	0	1
Tenure					
< 1	22367	0.16	0.36	0	1
1 - 2	22367	0.13	0.33	0	1
2 - 5	22367	0.27	0.44	0	1
5 - 10	22367	0.19	0.39	0	1
10+	22367	0.26	0.44	0	1
Occupational Group					
Management	22762	0.11	0.31	0	1
Professional	22762	0.12	0.32	0	1
Associate Professional	22762	0.17	0.37	0	1
Administrative	22762	0.19	0.39	0	1
Skilled Trade	22762	0.07	0.25	0	1
Personal Service	22762	0.09	0.28	0	1
Sales	22762	0.07	0.26	0	1
Machine Operatives	22762	0.08	0.26	0	1
Routine and Unskilled	22762	0.11	0.32	0	1
Union					
Yes	22329	0.37	0.48	0	1
No, but in the past	22329	0.17	0.37	0	1
No, never	22329	0.47	0.50	0	1
Acad. Qual.	21991	2.03	1.21	1	8
Voc. Qual	21022	1.30	0.71	1	9
Gender	22345	0.54	0.5	0	1
Weekly Hours Working	22114	35.93	12.45	0	96

Notes: This table provides information on the number of observations, mean and standard deviation as well as minimum and maximum values of control variables for estimations on income and discretion. Statistics for each variable are calculated omitting answers "refusal", "don't know" and "not applicable", all indicating unclear answers.

Table 7: Estimation of Income and Discretion Regressions

	Income	Discretion	
		What Tasks	How Work
	(1)	(2)	(3)
Occupational Group			
Professional	−8.12	0.33***	0.21***
	(8.87)	(0.028)	(0.023)
Associate	−85.3***	0.30***	0.19***
	(8.25)	(0.027)	(0.022)
Secretary	−158.1***	0.60***	0.33***
	(7.41)	(0.031)	(0.026)
Skilled Trade	−178.2***	0.54***	0.20***
	(8.37)	(0.042)	(0.031)
Personal Service	−221.2***	0.49***	0.35***
	(7.81)	(0.040)	(0.030)
Sales	−219.5***	0.65***	0.47***
	(7.87)	(0.043)	(0.034)
Operatives	−233.0***	0.79***	0.52***
	(8.74)	(0.042)	(0.037)
Unskilled	−264.2***	0.62***	0.37***
	(8.26)	(0.036)	(0.030)
Age			
Age 18 - 19	−8.09	−0.088	−0.065
	(12.7)	(0.093)	(0.077)
Age 20 - 21	10.4	−0.14	−0.11
	(12.8)	(0.094)	(0.077)
Age 22 - 29	35.3***	−0.15*	−0.13*
	(11.8)	(0.085)	(0.067)
Age 30 - 39	94.9***	−0.28***	−0.24***
	(11.5)	(0.085)	(0.067)
Age 40 - 49	106.6***	−0.30***	−0.24***
	(11.9)	(0.086)	(0.068)
Age 50 - 59	102.4***	−0.30***	−0.24***
	(11.9)	(0.087)	(0.067)
Age 60 - 64	72.7***	−0.38***	−0.34***
	(13.1)	(0.097)	(0.075)
Age 64+	16.2	−0.43***	−0.26**
	(21.2)	(0.13)	(0.11)

Tenure			
Tenure 1 - 2 years	4.88	-0.091^{***}	-0.00073
	(4.89)	(0.032)	(0.027)
Tenure 2 - 5 years	22.7^{***}	-0.12^{***}	-0.056^{**}
	(4.34)	(0.029)	(0.023)
Tenure 5 - 10 years	21.8^{***}	-0.17^{***}	-0.095^{***}
	(4.79)	(0.032)	(0.026)
Tenure $>$ 10 years	43.3^{***}	-0.25^{***}	-0.14^{***}
	(5.23)	(0.031)	(0.025)
Gender	-78.0^{***}	0.0049	-0.041^{**}
	(3.82)	(0.019)	(0.016)
Constant	160.0^{***}	2.30^{***}	2.01^{***}
	(16.4)	(0.098)	(0.078)
Employee Controls	*Yes*	*Yes*	*Yes*
Observations	21506	21420	21428
R^2		0.087	0.059

Notes: This table provides estimation results for interval regressions of income (column (1)) and linear regression of discretion on employee observables. Predictions from these regressions are used to generate estimated values for income and discretion for each employee.
Level of Significance: * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

A.2. Competency Tests

Table 8: Regression of Lab. Prod. on HRM-Cluster – Competency Tests

	What Tasks		How Work	
	(1)	(2)	(3)	(4)
HRM Cluster				
[0,0,0]		−0.31**		−0.33***
		(0.14)		(0.13)
[1,0,0]	−0.076	−0.39***	0.087	−0.25**
	(0.13)	(0.12)	(0.11)	(0.12)
[0,1,0]	0.097	−0.22	0.14	−0.19
	(0.16)	(0.15)	(0.15)	(0.15)
[0,0,1]	0.092	−0.22*	0.35***	0.015
	(0.14)	(0.13)	(0.11)	(0.12)
[1,1,0]	0.081	−0.23*	0.37***	0.035
	(0.15)	(0.14)	(0.14)	(0.15)
[1,0,1]	0.034	−0.28**	0.15	−0.19
	(0.14)	(0.14)	(0.13)	(0.14)
[0,1,1]	−0.041	−0.36**	0.15	−0.19
	(0.15)	(0.14)	(0.13)	(0.14)
[1,1,1]	0.31**		0.33***	
	(0.14)		(0.13)	
Union	−0.050	−0.050	−0.046	−0.046
	(0.085)	(0.085)	(0.084)	(0.084)
Pub. Sector	−0.065	−0.065	−0.070	−0.070
	(0.12)	(0.12)	(0.11)	(0.11)
Foreign	0.65***	0.65***	0.65***	0.65***
	(0.15)	(0.15)	(0.15)	(0.15)
Constant	3.69***	4.01***	3.52***	3.86***
	(0.20)	(0.21)	(0.19)	(0.21)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1812	1812	1812	1812
R^2	0.094	0.094	0.106	0.106

Notes: This table provides linear regression coefficients and standard errors of labour productivity on HRM cluster using competency tests as screening device and controls. The first panel (column (1) and (2)) refers to answers on the question "What Tasks" as proxy for discretion, columns (3) and (4) use "How to Work". Columns (1) and (3) use cluster [0,0,0] as base category; column (2) and (4) omit cluster [1,1,1].
Level of Significance: * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

Table 9: Regressions of Lab. Prod. on HRM Complementarities – Competency Tests

	What Tasks		How Work	
	(1)	(2)	(3)	(4)
High Income (I)	0.11	0.097	0.19	0.14
	(0.17)	(0.16)	(0.16)	(0.15)
High Discretion (D)	0.13	0.092	0.39***	0.35***
	(0.14)	(0.14)	(0.12)	(0.11)
Comp. Test (C)	−0.083	−0.076	0.076	0.087
	(0.13)	(0.13)	(0.12)	(0.11)
I × D	−0.17	−0.23	−0.36*	−0.35*
	(0.21)	(0.20)	(0.20)	(0.19)
I × C	0.097	0.061	0.038	0.14
	(0.21)	(0.20)	(0.21)	(0.19)
D × C	0.045	0.018	−0.26	−0.29*
	(0.18)	(0.18)	(0.18)	(0.17)
I × D × C	0.18	0.35	0.32	0.25
	(0.28)	(0.27)	(0.28)	(0.27)
Union		−0.050		−0.046
		(0.085)		(0.084)
Pub. Sector		−0.065		−0.070
		(0.12)		(0.11)
Foreign		0.65***		0.65***
		(0.15)		(0.15)
Constant	3.46***	3.69***	3.33***	3.52***
	(0.11)	(0.20)	(0.095)	(0.19)
Firm Controls	*No*	*Yes*	*No*	*Yes*
Subpop. Observations	1815	1812	1815	1812
R^2	0.017	0.094	0.035	0.106

Notes: This table provides linear regression coefficients and standard errors of labour productivity on binary variables of income, discretion, competency tests and its interactions. The first panel (column (1) and (2)) refers to answers on the question "What Tasks" as proxy for discretion, columns (3) and (4) use "How to Work".
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

A.3. Robustness

Table 10: Labour Productivity on HRM-Cluster: Robustness I

	Raw		Categorical		Log Income	
	"What" (1)	"How" (2)	"What" (3)	"How" (4)	"What" (5)	"How" (6)
HRM Cluster						
[0,0,0]	−0.37**	−0.40***	−0.38***	−0.53***	−0.38***	−0.50***
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
[1,0,0]	−0.27	−0.26	−0.31	−0.38**	−0.39*	−0.41**
	(0.22)	(0.20)	(0.20)	(0.19)	(0.21)	(0.21)
[0,1,0]	−0.36**	−0.29**	−0.22	−0.33**	−0.36**	−0.35**
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
[0,0,1]	−0.22	−0.23	−0.30**	−0.33**	−0.35**	−0.32**
	(0.15)	(0.15)	(0.14)	(0.15)	(0.14)	(0.15)
[1,1,0]	−0.66***	−0.64***	−0.43*	−0.61**	−0.43**	−0.53**
	(0.23)	(0.24)	(0.25)	(0.24)	(0.22)	(0.22)
[1,0,1]	−0.054	−0.10	−0.14	−0.24	−0.24	−0.29
	(0.17)	(0.19)	(0.18)	(0.19)	(0.17)	(0.18)
[0,1,1]	−0.23	−0.30**	−0.19	−0.28*	−0.27*	−0.32**
	(0.14)	(0.14)	(0.15)	(0.15)	(0.14)	(0.15)
Union	−0.048	−0.048	−0.053	−0.062	−0.058	−0.048
	(0.086)	(0.084)	(0.085)	(0.084)	(0.084)	(0.083)
Pub. Sector	−0.073	−0.063	−0.085	−0.070	−0.086	−0.058
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Foreign	0.58***	0.56***	0.56***	0.54***	0.57***	0.56***
	(0.14)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)
Constant	3.97***	4.00***	4.01***	4.07***	4.07***	4.08***
	(0.22)	(0.23)	(0.23)	(0.23)	(0.22)	(0.23)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1498	1498	1495	1495	1494	1494
R^2	0.100	0.099	0.091	0.103	0.088	0.095

Notes: This table provides linear regression coefficients and standard errors of labour productivity on HRM cluster and controls for different control specifications. Uneven columns refer to answers on the question "What Tasks" as proxy for discretion, columns (2), (4) and (6) use "How to Work".
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 11: Labour Productivity on HRM Cluster: Robustness II

	Aggregation		Only Pers. Tests		Only Comp. Tests	
	"What"	"How"	"What"	"How"	"What"	"How"
	(1)	(2)	(3)	(4)	(5)	(6)
HRM Cluster						
[0,0,0]	−0.41***	−0.43***	−0.50***	−0.50***	−0.31**	−0.33***
	(0.16)	(0.13)	(0.14)	(0.16)	(0.14)	(0.13)
[1,0,0]	−0.29	−0.31*	−0.40**	−0.36*	−0.39***	−0.25**
	(0.19)	(0.17)	(0.18)	(0.20)	(0.12)	(0.12)
[0,1,0]	−0.30*	−0.39***	−0.29**	−0.20	−0.22	−0.19
	(0.16)	(0.13)	(0.14)	(0.17)	(0.15)	(0.15)
[0,0,1]	−0.33**	−0.40***	−0.37***	−0.24	−0.22*	0.015
	(0.15)	(0.13)	(0.14)	(0.16)	(0.13)	(0.12)
[1,1,0]	−0.56**	−0.65***	−0.57**	−0.48*	−0.23*	0.035
	(0.24)	(0.22)	(0.26)	(0.28)	(0.14)	(0.15)
[1,0,1]	−0.23	−0.34*	−0.29*	−0.17	−0.28**	−0.19
	(0.20)	(0.18)	(0.17)	(0.19)	(0.14)	(0.14)
[0,1,1]	−0.31**	−0.33**	−0.37***	−0.33**	−0.36**	−0.19
	(0.15)	(0.14)	(0.14)	(0.16)	(0.14)	(0.14)
Union	−0.068	−0.061	−0.064	−0.042	−0.050	−0.046
	(0.084)	(0.084)	(0.084)	(0.084)	(0.085)	(0.084)
Pub. Sector	−0.078	−0.074	−0.068	−0.060	−0.065	−0.070
	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.11)
Foreign	0.57***	0.55***	0.55***	0.64***	0.65***	0.65***
	(0.13)	(0.13)	(0.12)	(0.13)	(0.15)	(0.15)
Constant	4.07***	4.12***	4.11***	4.02***	4.01***	3.86***
	(0.23)	(0.22)	(0.22)	(0.24)	(0.21)	(0.21)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1494	1494	1494	1494	1494	1494
R^2	0.089	0.091	0.093	0.103	0.094	0.106

Notes: This table provides linear regression coefficients and standard errors of labour productivity on HRM cluster and controls for different control specifications. Uneven columns refer to answers on the question "What Tasks" as proxy for discretion, columns (2), (4) and (6) use "How to Work".
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 12: Labour Productivity on HRM Complementarities: Robustness I

	Raw		Categorical		Log Income	
	"What"	"How"	"What"	"How"	"What"	"How"
	(1)	(2)	(3)	(4)	(5)	(6)
High Income (Inc.)	0.015	0.11	0.16	0.20*	0.025	0.15
	(0.11)	(0.10)	(0.099)	(0.10)	(0.11)	(0.10)
High Discretion (Disc.)	0.16	0.17*	0.081	0.20**	0.029	0.18*
	(0.11)	(0.10)	(0.099)	(0.094)	(0.11)	(0.095)
Pers. Test (PT)	0.11	0.15	0.073	0.14	−0.0065	0.091
	(0.19)	(0.16)	(0.17)	(0.15)	(0.18)	(0.17)
Inc. × Disc.	−0.027	−0.18	−0.053	−0.15	0.064	−0.15
	(0.15)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
Inc. × PT	−0.41	−0.49*	−0.28	−0.43	−0.068	−0.27
	(0.28)	(0.26)	(0.28)	(0.26)	(0.27)	(0.26)
Disc. × PT	0.057	−0.021	0.090	−0.049	0.12	−0.064
	(0.23)	(0.22)	(0.22)	(0.21)	(0.22)	(0.22)
Inc. × Disc. × PT	0.47	0.67**	0.31	0.61*	0.22	0.57*
	(0.33)	(0.33)	(0.35)	(0.33)	(0.32)	(0.33)
Union	−0.048	−0.048	−0.053	−0.062	−0.058	−0.048
	(0.086)	(0.084)	(0.085)	(0.084)	(0.084)	(0.083)
Pub. Sector	−0.073	−0.063	−0.085	−0.070	−0.086	−0.058
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Foreign	0.58***	0.56***	0.56***	0.54***	0.57***	0.56***
	(0.14)	(0.13)	(0.14)	(0.13)	(0.13)	(0.13)
Constant	3.60***	3.60***	3.62***	3.54***	3.68***	3.58***
	(0.21)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1498	1498	1495	1495	1494	1494
R^2	0.100	0.099	0.091	0.103	0.088	0.095

Notes: This table provides linear regression coefficients and standard errors of labour productivity on binary variables of income, discretion, personality tests and its interactions. Uneven columns refer to answers on the question "What Tasks" as proxy for discretion, columns (2), (4) and (6) use "How to Work".
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 13: Labour Productivity on HRM Complementarities: Robustness II

	Aggregation		Only Pers. Tests		Only Comp. Tests	
	"What"	"How"	"What"	"How"	"What"	"How"
	(1)	(2)	(3)	(4)	(5)	(6)
High Income (Inc.)	0.10	0.037	0.21**	0.30***	0.095	0.14
	(0.10)	(0.090)	(0.10)	(0.10)	(0.16)	(0.15)
High Discretion (Disc.)	0.082	0.035	0.13	0.25***	0.096	0.34***
	(0.10)	(0.10)	(0.11)	(0.092)	(0.14)	(0.11)
Pers. Test (PT)	0.12	0.12	0.032	0.11		
	(0.14)	(0.14)	(0.31)	(0.28)		
Comp. Test (CT)					−0.13	0.064
					(0.13)	(0.12)
Inc. × Disc.	−0.092	0.030	−0.21	−0.38***	−0.23	−0.34*
	(0.15)	(0.15)	(0.14)	(0.15)	(0.20)	(0.19)
Inc. × PT	−0.38	−0.38	−0.34	−0.42		
	(0.26)	(0.25)	(0.46)	(0.44)		
Inc. × CT					0.17	0.21
					(0.20)	(0.20)
Disc. × PT	−0.027	−0.065	0.18	0.073		
	(0.22)	(0.22)	(0.38)	(0.36)		
Disc. × CT					0.073	−0.29
					(0.20)	(0.20)
Inc. × Disc. × PT	0.59*	0.65**	0.75	0.84		
	(0.34)	(0.33)	(0.54)	(0.53)		
Inc. × Disc. × CT					0.27	0.22
					(0.29)	(0.29)
Union	−0.068	−0.061	−0.066	−0.047	−0.044	−0.041
	(0.084)	(0.084)	(0.089)	(0.090)	(0.091)	(0.091)
Pub. Sector	−0.078	−0.074	−0.088	−0.083	−0.087	−0.097
	(0.11)	(0.11)	(0.13)	(0.12)	(0.13)	(0.12)
Foreign	0.57***	0.55***	0.51***	0.61***	0.64***	0.66***
	(0.13)	(0.13)	(0.12)	(0.13)	(0.16)	(0.16)
Constant	3.66***	3.69***	3.62***	3.53***	3.70***	3.53***
	(0.20)	(0.21)	(0.22)	(0.21)	(0.21)	(0.21)
Firm Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Subpop. Observations	1494	1494	1108	1108	1108	1108
R^2	0.089	0.091	0.099	0.109	0.098	0.110

Notes: This table provides linear regression coefficients and standard errors of labour productivity on binary variables of income, discretion, personality tests (competency tests for columns (5) and (6)) and its interactions. Uneven columns refer to answers on the question "What Tasks" as proxy for discretion, columns (2), (4) and (6) use "How to Work".
Level of Significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

B. Figures

Figure 4: Distribution of Employee Questionnaires per Firm

This figure provides relative frequencies of returned questionnaires per firm. Only firms with a minimum of one questionnaires are included.

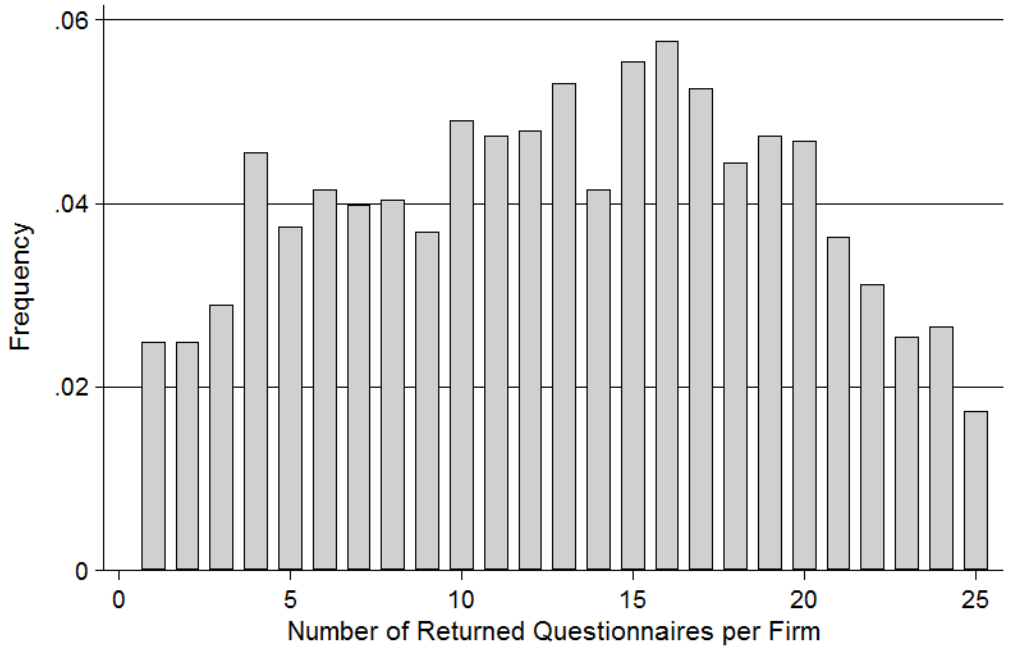
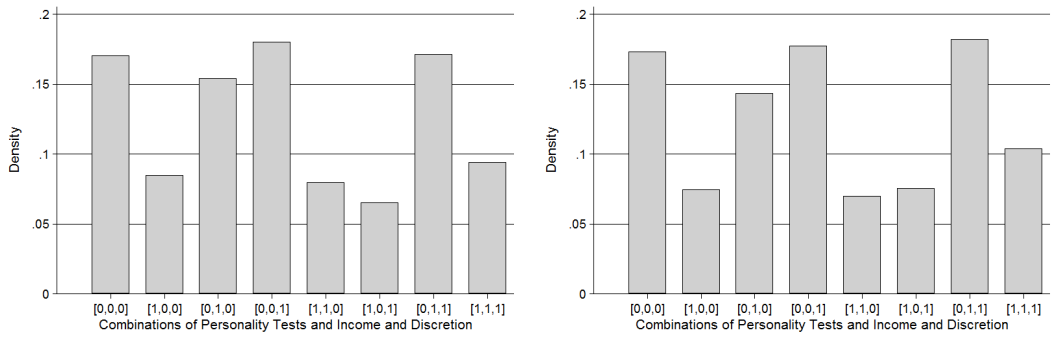


Figure 5: Distribution of HRM Clusters

This figure provides relative frequencies of HRM clusters, which were used in Section 4. Panel (a) refers to "What Tasks" as measure of discretion, panel (b) to "How to Work".



(a) Discretion: "What Tasks"

(b) Discretion: "How to Work"

Figure 5 depicts relative frequencies of each of the eight HRM clusters for two measures of discretion. The notation for each cluster is described as follows: [personality tests, high income, high discretion]. Each position is either 0 or 1 depending on whether firms require personality tests for job candidates, pays high income or allows for discretion. Hence on the very left of the figure cluster [0,0,0] describes firms which do not screen for personality, pay low income and do not grant high discretion and on the other extreme (cluster [1,1,1]) describes firms with personality tests, high income and discretion. Both panels exhibit similar frequency distributions. About 18 percent of firms belong to cluster [0,0,1], implying that these establishments do not screen job candidates for personality, do not pay higher than expected wages but provide substantial discretion. Approximately 10 percent of firms offer "good" jobs and screen for personality.