Maastricht University

You get what you pay for: Incentives and Selection in the Education System

Thomas Dohmen
Armin Falk

ROA Research Memorandum

You get what you pay for: Incentives and Selection in the Education System

Thomas Dohmen
Armin Falk

Abstract

You get what you pay for: Incentives and Selection in the Education System**

We analyse worker self-selection with a special focus on teachers. The point of the paper is that worker composition is generally endogenous, due to worker self-selection. In a first step we analyse lab experimental data to provide causal evidence on particular sorting patterns. This evidence sets the stage for our field data analysis, which focuses specifically on selection patterns of teachers. We find that teachers are more risk averse than employees in other professions, which indicates that relatively risk averse individuals sort into teaching occupations under the current system. Using survey measures on trust and reciprocity we also find that teachers trust more and are less negatively reciprocal than other employees. Finally, we establish differences in personality based on the Big Five concept.

Thomas Dohmen
ROA, IZA, DIW
Maastricht University
P.O. Box 616
6200 MD Maastricht
The Netherlands
t.dohmen@maastrichtuniversity.nl

Armin Falk
University of Bonn, IZA, CESifo and CEPR
Adenauerallee 24-42
D-53113 Bonn
Germany
emp@uni-bonn.de

In this paper we analyse the impact of incentives on worker self-selection, with a particular focus on the education system. Typically economists are interested in how incentives affect behaviour of a given workforce. The rationale for providing performance related pay is to align the interests of principals and agents in order to increase performance motivation. However, output does not only depend on effort provision but also on the composition of the workforce, i.e., employees' abilities, attitudes and personalities. The point of this paper is that worker composition is endogenous, due to worker self-selection: Agents with different characteristics and personality feel attracted by different types of incentives. In this sense providing incentives in firms or organisations has two important effects, an incentive effect per se and a selection effect. Importantly, these effects need not to be complementary. In the education system, for example, it may well be that changing from fixed pay systems to performance related pay schemes increases effort, but at the cost that currently selected teachers are dissatisfied and that the personality profile of newly attracted teachers fits requirements less well than under the status quo. On the other hand, if variable pay attracts more productive and able teachers, the selection effect positively adds to the incentive effect in terms of raising overall output. In this paper we do not study whether selection effects induced by performance incentives will have adverse or beneficial effects. What the paper shows, however, is that in as much as individual characteristics of teachers, and employees in general, matter for the success of organisations, neglecting selection effects is problematic.

We proceed in three steps. In the first step we discuss and reanalyse data from a laboratory experiment by Dohmen and Falk (2010).[1] The laboratory evidence sets

---

[1] Dohmen and Falk (2010) builds on data from additional sessions. The analyses in the current paper are based on the data that was available for the working paper version (see Dohmen and Falk, 2006)

the stage for our later analysis of sorting effects for teachers. The idea is to establish causality of sorting patterns in a controlled environment and to use these patterns to predict similar outcomes in the field. In the experiment subjects face the alternative to earn money either under a fixed wage regime or a variable pay scheme. We show that variable pay schemes attract more productive workers, as predicted by economic theory. We also show that women are less likely to choose a variable pay scheme and that being risk averse reduces individuals' willingness to work for variable pay. The advantage of laboratory data is that it provides the researcher with knowledge about individual characteristics while being able to observe the selection decision in a well-defined environment. This allows for drawing causal inferences, which is typically disputable when relying solely on field data.[2]

In the second step we analyse employee selection between the private and the public sector in Germany because teachers are public employees. The public sector is largely characterized by fixed payment schemes, whereas pay for performance schemes are more frequent in the private sector. In addition the threat of unemployment is basically absent in the public sector but provides strong performance incentives in the private sector. Based on the results from the laboratory experiment we predict that employees in the public sector should be more risk averse and that men are less likely to work in the public sector than women, compared to the private sector. These predictions are borne out by the data from the German Socio-Economic Panel Study (SOEP). In addition we find differences in personality and social preferences.

---

[2] For example, in the field agents are typically working under a mix of incentives such as explicit performance incentives and implicit contracts, rendering the isolation of incentive specific effects difficult. Moreover, one cannot rule out the possibility that changes in incentive schemes are endogenous, which renders causal inferences difficult. On the complementarity of lab and field experimental evidence see Falk and Heckman (2009).

In the third step we look specifically at selection patterns of teachers. We compare teachers with other similarly qualified employees. We find that teachers are more risk averse than employees in other professions, which indicates that relatively risk averse individuals sort into teaching occupations under the current system. Using survey measures on trust and reciprocity we also find that teachers trust more and are less negatively reciprocal than other employees. Finally, we establish differences in personality based on the Big Five concept.

Our results suggest that changes in the incentive system will affect well-being of incumbent teachers in the short-run and will change the composition of teachers in the long-run. Well-being is affected because variable pay exposes teachers to additional risk. Since incumbent teachers have selected on the basis of fixed payment schemes, job satisfaction is likely to fall among these relatively risk averse teachers. This effect is reinforced if the introduction of variable pay comes along with reduced job security. The composition is likely to change in the long run simply because the new remuneration scheme attracts different types of teachers. Whether the individual characteristics of the newly selected teachers are better or worse for overall outcomes in the education system cannot be answered with our data. It is likely that the effect on work attitude and ability is positive but that there are also potentially negative effects with respect to social preferences and teacher personality in general.

Our findings complement an extensive literature on teacher supply. This literature has focused mainly on the role of relative wages, earnings profiles or career considerations (for an overview see Dolton, 2006). In contrast we concentrate on the role of incentive pay and the interaction of personality and modes of payment. This is particularly relevant given the importance of teacher quality for student achievement and the fact that non-pecuniary factors such as personality must play an important role

to become a teacher at all. This follows from Dolton (2006), who, summarizing evidence for the USA and the UK, concludes that "if earnings were the only criteria used to decide a career then no men would become teachers in these countries." p. 1150).

The rest of the paper is organized as follows. Section 2 briefly summarizes arguments for why personal characteristics such as ability, risk aversion, social preferences and gender might affect the decision to sort into particular pay schemes. Section 3 discusses the laboratory evidence on worker self-selection and sets the stage for our analysis of employees in the private vs. the public sector as well as teachers vs. other employees (Section 4). Section 5 concludes.

1. Theoretical Arguments for the Effects of Remuneration Schemes on Sorting

If subjects face the choice between a fixed wage contract and a variable pay contract, it is obvious that the probability of preferring the variable pay contract is higher the more productive a worker is. To exemplify, consider the choice between a fixed-payment contract that pays a wage of w_{fixed} independent of output and a piece rate contract that pays a piece rate of β for each unit of output x produced, so that the wage w under the piece rate scheme equals βx. Furthermore assume that output is produced according to the production function $x = \pi(\theta, e) + \epsilon$, where θ measures ability, e measures costly effort and ϵ is a white noise error term. If a worker's utility depends positively on the wage w and negatively on effort e according to the utility function $u(w, e) = w - c(e)$ with $u'(c) < 0$ and $u''(c) > 0$, it is straightforward to show that a worker whose productivity exceeds a certain threshold value optimally opts for the piece-rate contract, while a worker with lower productivity prefers the fixed-payment. In the absence of risk (i.e., without random noise in production) or in

case of risk neutrality, this productivity threshold is given by $\hat{\pi} = \frac{w_{fixed} + c(e^*) - c(e_{min})}{\beta}$, where e^* is the optimal effort level that satisfies $\frac{\delta c(e)}{\delta e} = \beta \frac{\delta \pi}{\delta e}$ and e_{min} is the minimal enforceable effort level in the fixed wage regime. The productivity threshold increases in the level of the fixed payment, and it decreases in the attractiveness of the piece rate. In a similar vein it can be shown that more productive workers are also more likely prefer other prototypical variable pay schemes such as tournament or revenue-sharing contracts to fixed wage contracts.

It is also clear that risk aversion affects the choice between variable pay and fixed wage contracts. Principal-agent theory has emphasized that risk-averse workers dislike the income risk that is associated with variable pay when output depends upon factors beyond their control, e.g., if random noise affects production. In such a situation, income under variable pay schemes are risky, while no such risk is eminent in a fixed wage regime in which payments are independent of output. As a consequence the expected utility from variable pay is lower for risk averse subjects than for risk neutral or risk loving subjects. Hence we expect that subjects are less likely to select into variable pay the more risk averse they are.

Another potential driver of self-selection are social preferences. Traditional contract theory is based on the assumption that principals and agents are solely interested in their own material payoffs. However, considerable evidence indicates that a substantial fraction of people also care about reciprocal fairness (see, e.g., Fehr and Gächter, 2006). The co-existence of selfish and reciprocally motivated agents changes the optimality conditions of different types of contracts (see Fehr, Klein and Schmidt, 2007). Recently, several theoretical studies have highlighted how social preferences affect optimal contract design (see, e.g., Grund and Sliwka, 2005; Englmaier and Wambach, forthcoming; and Kosfeld and von Siemens, 2007).

Finally, several studies have emphasized that gender affects the sorting decision into competitive compensation schemes (e.g., Niederle and Vesterlund, 2007). These studies suggest that women are more likely than men to prefer non-competitive and non-variable pay.

2. Self-selection in the Laboratory

In this section we discuss the experiment by Dohmen und Falk (2010) and reanalyse their data. The idea of their experiment is to show the existence of multidimensional sorting, i.e., to isolate individual characteristics that are relevant in the process of sorting into different incentive schemes. While many laboratory experiments have studied the incentive effects of different incentive systems, such as piece-rate or tournaments, only few experiments have systematically investigated whether and how agents with different characteristics feel attracted by different incentives schemes (see Dohmen and Falk (2010) for a discussion).

Dohmen and Falk (2010) study self-selection into variable pay focusing on the role of individual productivity, willingness to take risks, social preferences, relative self-assessment, personality and gender. The experiment consists of three phases, (i) elicitation of productivity, (ii) sorting and working under different incentives and (iii) elicitation of personal characteristics. The work task consists of the multiplication of a two-digit and a one-digit number with five degrees of difficulty.[3] Difficulty varies because different levels of working memory are involved. This real effort task is useful for the purpose at hand, as it requires no previous knowledge and is simple to explain. Moreover, the task sufficiently differentiates subjects' performance, and task

---

[3] Examples for the different difficulty levels are: Level 1: 11*9 = ???; Level 2: 3*32 = ???; Level 3: 6*43 = ???; Level 4: 4*68 = ???; Level 5: 7*89 = ???

performance is a proxy for general cognitive ability. Learning effects that could possibly confound the measure are expected to be absent.

In the first phase of the experiment we gauge individual productivity. One of the indicators measures how fast subjects can solve a multiplication problem. To this end they are asked to calculate a multiplication problem of difficulty level 4 as fast as possible. The time that elapsed before the correct answer was entered on the computer screen is used as an indication for productivity related to the specific task of the experiment. It turns out that this indicator is also a pretty good proxy for general ability as it is significantly correlated with the final math grade in high school — a measure that ranges from 1, the worst grade, to 15, the best grade — and with the final high-school grade point average (Abitur) — a measure that ranges from 4.0, the worst grade, to 1.0, the best grade. The Spearman rank correlations and corresponding p-values of math grades and Abitur grades are -0.28 ($p < 0.001$) and 0.29 ($p < 0.001$), respectively. Later we will use the productivity indicator together with grades to study ability based self-selection.

After the elicitation of individual productivity indicators subjects made a decision how to earn money in a 10 minute working phase. They could either choose to work for a fixed payment or a variable payment. The fixed payment guaranteed an amount of money (400 points) independent of correctly solved multiplication problems. In contrast, if a subject chose a variable payment scheme, actual payments varied in the number of correctly solved problems. Dohmen and Falk (2010) study three treatments. In each treatment subjects can choose between a fixed payment and a variable payment. The difference between treatments is the type of variable pay: piece rate, tournament or revenue sharing. In the piece-rate condition subjects receive a piece rate (10 points) for each correctly solved problem. In the tournament treatment

two subjects compete against each other. The subject with the higher number of correctly solved problems gets a high prize (1300 points), the other gets a low prize (0 points). Finally, in the revenue-sharing condition two subjects are paid according to their joint output. Joint output is multiplied by 10 (as in the piece-rate treatment) and equally divided between the two members of the team.

In the third phase of the experiment, we elicited several preference and personality indicators. In this paper we will concentrate on the role of risk attitudes, which we measured in the experiment with a simple questionnaire measure. We use answers on an 11-point scale to the question "*How do you see yourself: Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?*", where the value 0 means '*completely unwilling to take risks*' and the value 10 means '*completely willing to take risks'*. We use this question for two reasons. First, it is exactly the same as we use below to study risk attitudes in the SOEP. This eases comparison between our lab and field evidence. Second, we are confident about the behavioral validity of this survey measure, because it predicts a wide variety of economically relevant behaviors (e.g., investment in stocks, self-employment, and migration behavior) and has been validated in a large-scale field experiment, using a representative sample of 450 German adults (for details see Dohmen et al., 2010).

In this paper we pool the data from all three treatments and study how productivity and other individual characteristics affect the likelihood to self-select into variable pay. In total 56 percent of the 240 subjects chose the variable pay scheme.[4] Table 1 shows how these subjects can be characterised. The first column of Table 1 shows that productivity has a positive effect on working under variable pay. We use two measures of individual productivity or ability, the time it takes to solve a

---

[4] In order to have an even sample of men and women, we recruited in a way such that out of our 240 subjects 121 are female and 119 are male. In Dohmen and Falk (2010) we ran additional sessions and analyse decisions of 360 subjects.

problem of difficulty level 4 (Productivity) and the final high-school grade point average (Abitur), which ranges from 1.0 (very good) to 4.0 (sufficient). We also include dummies for the piece-rate treatment (PT) and the revenue sharing treatment (RT). The tournament treatment is the reference category. The coefficient estimates for *Productivity* and *Abitur* indicate that subjects who are faster in solving a problem and have better grades are significantly more likely to self-select into the variable pay schemes. This sorting pattern is intuitive and exactly what economic theory would predict.[5] Subjects who are slow or have relatively high costs of solving the problems are better off choosing the fixed payment, while highly productive subjects can earn much more in the variable than the fixed payment schemes.

Interestingly women are significantly less likely to sort into variable payment schemes than male subjects. This effect is large; women are about 20 percent less likely to select into variable pay. It suggests that, controlling for productivity[6], women seem to dislike the uncertainty and/or competitiveness that is inherent to variable pay schemes. This gender specific sorting pattern is also found in experiments that analyse sorting decisions between variable pay schemes that vary the degree of competitiveness, e.g., piece-rate vs. tournament. In these experiments, women favour less competitive environments (see, e.g., Niederle and Vesterlund, 2007).

A major difference between fixed and variable pay is that the former is a risk-free option while the latter involves uncertainty and risk. Economic theory therefore predicts that the more risk tolerant an agent is the more likely he should self-select into the variable pay ceteris paribus. Results in Table 1 show that subjects who are relatively willing to take risks are significantly more likely to self-select into variable

---

[5] See also Cadsby et al. (2007), who report corroborative evidence on productivity sorting in which subjects had the choice of working on an anagram word-creating task.

[6] Note that women solve less problems than men. Including more detailed controls for productivity therefore reduces the gender effect (see Dohmen and Falk 2010).

pay, as predicted. As indicated by the marginal effects estimates, being one point more risk averse on the 11-point scale reduces the likelihood of selecting into variable pay by about 5.5 percentage points. This means that for a given productivity variable pay attracts relatively risk tolerant agents.[7]

In column 2 we also include personality indicators. Personality has not received much attention in economics in general (Borghans et al., 2008) and it is therefore not surprising that little is known about the relation between personality and sorting into different payment schemes or firms. This is a real shortcoming because personality probably matters to firms as much as other more standard personal characteristics, such as formal qualifications or ability. In fact, employer surveys suggest that so-called "soft skills" such as reliability or positive work attitudes are rated by employers as more important than prior work experience or technical skills (Regenstein, Meyer, and Hicks, 1998; Becci et al., 2005; Atkinson and Williams, 2003). This explains why firms make use of personality tests in their hiring process (see, e.g., Autor and Scarborough, 2005). We believe that personality is also relevant in the education system, given the close and intense personal interaction between teachers and students. Our personality measure was developed by Hermann Brandstätter (see Brandstätter, 1988).[8] Column 2 shows that personality sorting is not particular pronounced in our pooled data set. Only the personality trait "ability to

---

[7] It is interesting to note that the gender coefficient is larger without controls for risk attitudes. This is intuitive since women are generally less willing to take risks than men. This holds in our sample of 240 subjects as well as in a large representative sample from the SOEP (see Dohmen et al., 2010). Thus at least in part gender specific sorting can be explained by gender specific risk attitudes.

[8] This so called 16 PA test is a short form of the German-language version of Cattell's sixteen personality factor questionnaire (16 PF), developed by Schneewind, Schröder and Cattell (1983). Sixteen primary factors of personality are constructed based on 192 items, and can be summarized in five independent secondary factors. The Brandstätter test presents subjects with 32 conflictive adjective pairs. For each adjective pair, subjects rate how well an adjective pair describes their personality on a 9-point scale. Linear combinations of these 32 ratings produce the five secondary factors, following a procedure described by Brandstätter (1988). The coefficients of these linear combinations were determined in a regression analysis, in which each of the five secondary factors that were obtained from the 16 PF-test of Schneewind, Schröder and Cattell (1983) was regressed on all 32 items of the 16 PA test for a sample of 300 individuals who had completed both tests.

work under pressure" is weakly significant, indicating that people who are relatively able to work under pressure have a higher tendency to self-select into variable pay. The fact that other traits are not significant could imply that personality as measured by the Brandstätter questionnaire is not strongly related to selection decisions into variable pay.[9]

Table 1

To sum up our data suggest that incentives systematically affect sorting in multiple dimensions. Variable pay schemes systematically attract people with different abilities and individual characteristics than fixed pay contracts do. In particular, more productive subjects sort into the variable pay schemes, while relatively risk averse subjects are less likely to self-select into variable pay. Women are less likely to sort into variable payment schemes, an effect that seems to be driven in part by risk preferences. It is noteworthy that on top of the selection patterns we also find a straightforward incentive effect. According to self-reported measures of work effort, stress and exhaustion we find that subjects in variable pay schemes provide significantly more effort. They also feel more stressed and exhausted than subjects working for the fixed payment (for details see Dohmen and Falk 2010).

3. Selection in the Field

Building on our laboratory findings, in this section we move from the lab to the field. We first study differences in individual characteristics of employees in the private vs. the public sector before we explicitly study self-selection of teachers. In Germany, teachers are employed in the public sector, which is largely characterized by fixed payment schemes. In addition the threat of unemployment is basically absent in the

---

[9] One should note, however, that we have pooled our data in Table 1. If we look at sorting decisions separately by different incentive schemes and gender, personality differences are more pronounced.

public sector, in particular for civil servants. In contrast the risk of layoff provides strong performance incentives in the private sector. We would therefore expect a similar sorting pattern as is shown in Table 1, both for public employees as well as teachers. Using data from the SOEP we run a similar regression as in Table 1 including measures for productivity, risk preferences, gender, social preferences and personality.

The SOEP is a representative panel survey of the adult population living in Germany. To ease comparison with our laboratory data we use school grades as productivity proxies. Risk attitudes are measured by the exact same measure as in Table 1. We also include measures for social preferences, as they are relevant for the design of optimal contracts (Fehr and Falk, 2002) but also because they could be relevant for teachers as we will discuss below. Social preferences are measured in terms of trust and reciprocity. The trust measure is the principal component of answers to three questions in the 2003 wave about individuals' trust attitudes (cf. Dohmen et al., 2008): *(1) In general, one can trust people. 2) These days you cannot rely on anybody else. 3) When dealing with strangers it is better to be careful before you trust them*. The four answer categories were labelled: *strongly agree, agree somewhat, disagree somewhat, strongly disagree*. Similar to the survey risk measure the trust measure has been validated in a paid experiment.[10] Reciprocity is measured as positive and negative reciprocity: The measures are obtained by averaging respondents' agreement (on a 7-point scale) to three statements each.[11] Agreement to

---

[10] Fehr et al. (2002) summarize an individual's responses to the three survey questions about trust using factor analysis, and then show that this combined measure is a significant predictor of the amount that a first-mover actually sends to the other player, in a paid trust game experiment.

[11] The measure for positive reciprocity is based on the following three statements: *If someone does me a favor, I am prepared to return it; I go out of my way to help somebody who has been kind to me before; I am ready to undergo personal costs to help somebody who helped me before.* The measure for negative reciprocity is based on the following three statements: *If I suffer a serious wrong, I will take revenge as soon as possible, no matter what the cost; if somebody puts me in a difficult position, I will do the same to him/her; if somebody insults me, I will insult him/her back.*

these statements is indicated on a scale from 1 (meaning "*does not apply to me at all*") to 7 ("*applies to me perfectly*").

We restrict the sample to employees who have completed a high school exam (Abitur) in order to ensure that we compare teachers to other employees with a similar educational background: A formal requirement for teachers in Germany is to complete university study. The Abitur is a prerequisite for entering university. Our sample consists only of individuals who are not older than 65 and are currently neither enrolled in school, nor in vocational training, nor in university. Furthermore, we drop observations with item non-response on the high-school grades in math and German, gender, age, risk attitude, as well as the trust and reciprocity measures. This leaves us with a sample for analysis that covers 1521 individuals among whom 608 work in the public sector and 287 are teachers (high school teachers, primary school teachers, university professors and teachers in various other education institutions).

Results of Probit estimates, in which the dependent variable is an indicator variable for working in the public sector, are shown in Columns (1) and (2) of Table 2.[12] Column 1 shows a weak and insignificant productivity effect as measured by the average math and German grade at age 17. Women are significantly more likely to work in the public sector; and public sector employees are significantly less willing to take risks than private sector employees. Given the differences with respect to risk and uncertainty in the private vs. the public sector, this is exactly what we would

---

[12] In the following some caution in interpreting the selection effects is warranted. When comparing employees in the private and public sector (and teachers vs. other employees) we do not only compare differences in the incentive structures but also different jobs with job specific characteristics. While in the experiment we kept the task constant while changing incentives, we now compare job-incentive-bundles. In this sense some of the characteristics that we identify as being relevant for sorting may depend on specific job characteristics, not just on different incentive structures. Unlike with our experimental data it is also possible that in the SOEP data attitudes and preferences are endogenous to the work environment, rendering causal interpretations difficult.

expect. It also corroborates our laboratory findings.[13] We also detect significant differences with respect to social preferences. As shown in Column 1, employees in the public sector trust more than in the private sector. This effect, however, becomes insignificant once we include the Big Five personality indicators in Column 2.

Table 2

The Big Five is a wide spread scientific attempt to classify personality traits. Each of the Big Five factors comprises many different, specific characteristics of personality. The Big Five taxonomy serves an integrative function because it can represent the various and diverse systems of personality descriptions in a common framework. Although being the most used concept, it has also been criticised. For example, it has been pointed out that there is no theory why one should cluster personality reliably into 5 groups (and there are alternative suggestions using different numbers of factors). Moreover, the Big Five model suppresses important sub-factors, which might explain important individual differences, such as persistence, motivation, regulation of motivation etc. (compare Borghans et al., 2008). The 2005 wave of the SOEP contains a short version of Big Five, including 15 items, three for each factor (for details see Gerlitz and Schupp, 2005). Respondents are offered different adjectives and indicate whether they apply or not on a 7-point scale (compare Table 3).

Table 3

In Column (2) of Table 2 two factors turn out to be significant: public employees are less conscientious and more open to experience. High values in conscientiousness imply that a respondent "does a thorough job", does not "tend to be lazy" and "does things effectively and efficiently". Employees in the public sector are significantly less conscientious compared to those working in the private sector. The other factor,

---

[13] Additional complementary evidence is provided by Bonin et al. (2007), who find that individuals who are more willing to take risks are more likely to work in occupations with higher earnings variability.

openness to experience, is measured by the following items: "is original", "comes up with new ideas and values artistic experiences" and "has an active imagination". This factor is significantly higher for public employees.

We now turn to the selection profile of teachers and analyse how teachers differ from other employees, i.e., we compare teachers to all employees in the SOEP who have obtained a schooling degree that qualifies them for attending university (Abitur). We estimate Probit models in Columns (3) and (4) of Table 2 in which the dependent variable takes value 1 if the individual has self-selected into a teaching profession.

The results are similar to those that we obtain when comparing employees from the public with those in the private sector (see Columns (1) and (2)). The marginal effects estimates of the Probit model in Column (3) indicate that women are about 13 percent more likely to become teachers than men. In terms of personality and attitudes, there is a significantly negative correlation between willingness to take risks and the probability of selecting into the education sector. Under the current payment schemes, which are characterised by fixed salaries and job security, risk averse individuals are significantly more likely to become teachers. The coefficient indicates that for each point on the 11-point questionnaire scale, the likelihood of becoming a teacher changes by almost two percentage points. We also find significant effects concerning social preferences: teachers are significantly more trusting and less negatively reciprocal compared to the comparison group (see Column 3). In Column 4 we include personality indicators according to the Big Five model. It turns out that teachers are less "conscientious", indicating a relatively low work attitude.

Up to this point we have treated teachers as a homogeneous group. We think that this is justified to the extent that most teachers face a similar incentive structure,

all being employed in the public sector and having comparable job characteristics. There are, however, some notable differences. For example, high school teachers have some limited promotion opportunities, which are more or less absent for primary school teachers (see Wößmann, 2008). Also the job profile and the required qualifications differ substantially between school teachers and university professors. In order to investigate how differences in incentives and job profiles among teaching careers might trigger differential sorting patterns, we estimated a multinomial logit model in which the dependent variable comprises the reference group of non-teaching professions and five categories for different teaching subgroups: primary school teacher, junior high-school teacher (Hauptschule, Realschule), high-school (Gymnasium, Gesamtschule, Berufsschule), college and university and other teaching professions. The results, which are shown in Table 4, largely corroborate our findings from Table 2, but they also provide additional interesting insights. A comparison of primary school teachers and university professors illustrates this: First, and not surprisingly, the gender effect is different for university professors than for primary school teachers. While the latter are significantly more likely to be female, the former are more likely to be male. In terms of productivity as measured by grades we find that grades are better for university teachers and worse for primary school teachers. In terms of risk, reciprocity and trust most coefficients are similar and have the same sign, which holds interestingly for all teacher types. Note, however, that high school teachers are less risk averse than primary school teachers, which is consistent with differences in the promotion opportunities across these two sub-groups. In terms of personality, primary school teachers are significantly less "conscientious", while university professors are more "open to experience" and show less "extraversion".

To sum up we find systematic differences between employees in the public vs. the private sector as well as between teachers and those working in other occupations. Our laboratory findings suggest that systematic self-selection produces these observed differences. These selection effects are important when considering the consequences of introducing monetary incentives for employees in the education sector, for at least three reasons. First, changes in incentive schemes create potential tension and dissatisfaction. The reason is simply that employees in the education system, i.e., teachers, who have (optimally) self-selected into the education system based on the premise that remuneration is characterised by non-contingent pay would be confronted with payment schemes that make payments contingent on output and performance. If the selection decision was optimal, the new work environment will not be satisfying, i.e., for a given salary job satisfaction of teachers would drop. This effect will last until a newly selected sample of teachers will be in place. The reduction in utility may lead to resistance against pay for performance reforms, potentially leading to conflict and reduced effort and lower quality of teachers' work output. Second, and related to this first point, it may not be efficient to intensify monetary incentives if teachers are *genuinely* risk averse. A well-known insight from incentive theory is the existence of a trade-off between risk and incentives. Take a simple linear incentive scheme

$$w = \alpha + \beta x \beta \qquad (1)$$

where w is the wage, α is a basic compensation independent of output, β is the incentive intensity and x is output, which is a combination of effort and luck, i.e., $x = e + \varepsilon$, with e = effort and ε is an error term. Under the assumption of a risk neutral principal, the efficient incentive intensity β is

$$\beta^* = P'(e)/1 + rc''(e)Var(\varepsilon) \qquad (2)$$

where *P'(e)* is the marginal profit increase in effort, *r* is a measure of risk aversion of the agent, *c''(e)* measures the sensibility of effort reaction to a change in *b (de/db = 1/c''(e))* and *Var(ε)* measures the precision of measuring effort. Here we are interested in the relation between preferences and incentives. According to equation (2) the efficient level of *β* is lower, the higher the risk aversion of the agent. The intuition is that incentives expose agents to risks, for which they have to be compensated. Thus, if teachers are relatively risk averse, variable pay schemes are less efficient than if teachers were risk neutral.

Third, the results in Tables 2 and 4 suggest that people who show relatively high levels of trust and low levels of negative reciprocity feel less attracted by variable pay. In other words one would expect that introducing variable pay in the educational system may eventually crowd in teachers who are less trusting and more negatively reciprocal at the cost of the current teacher profile. We can only speculate that changing the composition of teachers along this dimension may adversely affect the educational production process. For example, teachers are important role models for students and affect attitude formation of their students. Evidence shows, for example, that social preferences such as trust are partly determined by environmental factors (Dohmen et al., 2006). Dohmen et al. (2009) use SOEP data and show that the less negatively reciprocal people are, the lower is their unemployment risk and the higher their subjective well-being. In terms of the role model argument it is therefore good news that teachers are significantly less negatively reciprocal. In this sense, there is the danger that introducing variable pay attracts the "wrong" teachers. On the other hand we have seen a mild effect of productivity and a systematic effect along the Big Five factor conscientiousness. Introducing variable pay may increase average productivity of teachers as well as the fraction of conscientious teachers, i.e., teachers

with positive work attitudes. We do not know whether the overall impact of changing incentives and thereby changing the composition of teachers is positive or negative; but we can say that the composition will be different and that this effect should be considered when considering the optimality of introducing incentives in the education system.

4. Concluding Remarks

In this paper we have argued that introducing incentives has two distinct effects, an incentive effect and a selection effect. In a laboratory experiment we have shown that fixed vs. variable pay schemes attract systematically different types of workers, along several dimensions. A similar multi-dimensional sorting pattern was also observed among employees in the public vs. private sector as well as among teachers and other employees. We have further argued that changing incentives will affect well-being of those employees who have self-selected into a particular job and, in the long-run change the composition of employees. These effects need to be accounted for when changing incentive schemes in organisations.

Incentive induced sorting effects are likely not only relevant for teachers, i.e., the supply side of education, but also for students, i.e., the demand side. Changing incentives that relate to the financing of the education system, e.g., by introducing tuition fees for students, alters incentives to invest in education and will therefore affect the composition of students. Since the costs of an investment in human capital accrue in the present (in the form of tuition fees, forgone earnings and psychic costs), while the private financial returns of such an investment manifest only in the future, the human capital model (Becker, 1975; Mincer, 1958 and 1974; Ben-Porath, 1967) predicts that a higher discount rate is associated with lower levels of educational

attainment (Hause, 1974). Likewise, uncertainty about the returns to education implies a negative relationship between an individual's degree of risk aversion and educational investment (Levhari and Weiss, 1974; Hartog and Diaz Serrano, 2007).[14] Exploring how risk and time preferences affect educational choice is therefore an important issue for future research. If educational investments depend on these preferences, it is of similar importance to assess how personal and socio-economic characteristics determine preference heterogeneity. For example, if higher parental wealth is associated with more willingness to take risks, individuals from wealthier families are more inclined to invest in schooling *ceteris paribus*, and policies that increase uncertainty about returns to education would predominately deter individuals from poorer households from studying. We believe that understanding how preferences and incentives interact to determine sorting decisions in the education sector is crucial for policy design.

References

Atkinson, J., and Williams, M. (2003). 'Employer perspectives on the recruitment, retention and advancement of low-pay, low-status employees'. UK: Government Chief Social Researcher's Office.

Autor, D.H. and Scarborough, D. (2005). *'Will job testing harm minority workers? Evidence from the retail sector'*. MIT, Department of Economics.

Becker, G. (1975). *Human Capital*. New York: Columbia University Press.

Ben-Porath, Y. (1967). 'The production of human capital and the life cycle earnings.' *Journal of Political Economy*, vol. 75, pp. 352-65.

Brandstätter, H. (1988). 'Sechzehn persönlichkeits-adjektivskalen (16 PA) als forschungsinstrument anstelle des 16 PF. *Zeitschrift für experimentelle und ange- wandte Psychologie*, vol. 35, no. 3, pp. 370-91.

---

[14] On the other hand, it has been argued that education can insure against macroeconomic shocks that affect, for example, the risk of future unemployment (Gould et al., 2001). From a theoretical perspective, it is therefore a priori not clear how risk preferences are related to human capital investments. Empirically, however, Dominitz and Manski (1996) find that individuals perceive education as risky, so that we expect individuals who are more willing to take risks to acquire more education.

Bonin, H., Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2007). 'Cross-sectional earnings risk and occupational sorting: the role of risk attitudes.' *Labour Economics*, vol. 14. no. 6, pp. 926-37.

Borghans, L., Duckworth, A.L., Heckman, J.J. and ter Weel, B. (2008). 'The economics and psychology of personality traits.' *Journal of Human Resources*, vol. 43, no. 4, pp. 972-1059.

Cadsby, B.C., Song, F. and Tapon, F. (2007). 'Sorting and incentive effects of pay-for-performance: an experimental investigation.' *The Academy of Management Journal*, vol. 50, no. 2, pp. 387-405.

Card, D. (1999). 'The causal effect of education on earnings'. In *Handbook of Labor Economics* (eds O. Ashenfelter and D. Card), vol. 3A. North-Holland: Elsevier Science.

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2006). 'The intergenerational transmission of risk and trust attitudes', Discussion Paper No. 2380, IZA.

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2008). 'Representative trust and reciprocity: prevalence and determinants.' *Economic Inquiry,* vol. 46, no. 1, pp. 84-90*.*

Dohmen, T., Falk, A., Huffman, D. and Sunde, U. (2009). 'Homo reciprocans: survey evidence on behavioral outcomes.' Economic Journal*,* vol. 119, no. 536, pp. 592-612.

Dohmen, T. and Falk, A. (2006). 'Performance pay and multi-dimensional sorting: productivity, preferences and gender. Discussion Paper No. 2001, IZA.

Dohmen, T. and Falk, A. (2010). 'Performance pay and multi-dimensional sorting: productivity, preferences and gender.'*American Economic Review*, forthcoming.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J. and Wagner, G.G. (2010). 'Individual risk attitudes: measurement, determinants and behavioral consequences,' *Journal of the European Economic Association*, forthcoming.

Dolton, P.J. (2006). 'Teacher supply'. In*Handbook of the Economics of Education* (eds E.A. Hanushek and F. Welch), vol. 2.

Dominitz, J. and Manski, C. (1996). 'Eliciting student expectations of the returns to schooling.' *Journal of Human Resources*, vol. 31, pp. 1–26.

Englmaier, F. and Wambach, A. (Forthcoming). 'Optimal incentive contracts under inequity aversion.' *Games and Economic Behavior*.

Falk, A. and Heckman, J.J. (2009). 'Lab experiments are a major source of knowledge in the social sciences.' *Science*, vol. 326, no. 5952, pp. 535-38.

Fehr, E. and Falk, A. (2002). 'Psychological foundations of incentives.' *European Economic Review*, vol. 46, pp. 687-724.

Fehr, E., Fischbacher, U., von Rosenbladt, B., Schupp, J. and Wagner, G.G. (2002). 'A nation-wide laboratory - examining trust and trustworthiness by integrating behavioral experiments into representative surveys. ' *Schmollers Jahrbuch*, vol. 122, no. 4, pp. 519-42.

Fehr, E. and Gächter, S. (2006). *Reciprocity and Contract Enforcement*. Amsterdam, North-Holland: Elsevier Science.

Fehr, E., Klein, A. and Schmidt, K.M. (2007). 'Fairness and contract design.' *Econometrica*, vol. 24, no. 1, pp. 121-51.

Gerlitz, J-Y. and Schupp, J. (2005). 'Zur Erhebung der Big-Five-basierten Persönlichkeitsmerktmale im SOEP.' DIW Research Notes, no. 4.

Golsteyn, B.H.H. (2007). *The Ability to Invest in Human capital.* Maastricht: ROA.

Gould, E.D., Moav, O. and Weinberg, B. (2001). 'Precautionary demand for education, inequality, and technological progress.' *Journal of Economic Growth*, vol. 6, no. 4, pp. 285-315.

Grund, C. and Sliwka, D. (2005). 'Envy and compassion in tournaments.' *Journal of Economics & Management Strategy*, vol. 14, no. 1, pp. 187-207.

Hartog, J. and Diaz-Serrano, L. (2007). 'Earnings risk and demand for higher education.' *Journal of Applied Economics*, vol. 10, no. 1, pp. 1-28.

Hause, J.C. (1974). 'The risk element in occupational and educational choices: comment.' *Journal of Political Economy*, vol. 82, no. 4, pp. 803-7.

Kosfeld, M. and von Siemens, F. (2007). 'Competition, cooperation, and corporate culture', Discussion Paper, no. 2927, IZA.

Lazear, E.P. (2000). 'Performance pay and productivity.' *American Economic Review*, vol. 90, no. 5, pp. 1346-62.

Levhari, D. and Weiss, Y. (1974). 'The effect of risk on the investment in human capital.' *American Economic Review*, vol. 64, no. 6, pp. 950-63.

Mincer, J. (1958). 'Investment in human capital and the personal income distribution.' *Journal of Political Economy*, vol. 66, pp. 281-302.

Mincer, J. (1974). *Schooling Experience and Earnings*. New York: Columbia University Press.

Munasinghe, L. and Sicherman, N. (2006). 'Why do dancers smoke? smoking, time preference, and wage dynamics.' *Eastern Economic Journal*, vol. 32, no. 4, pp. 595-616.

Newton, B., Hurstfield, J., Miller, L., Page, R. and Akroyd, K. (2005). 'What employers look for when recruiting unemployed and inactive: skills, characteristics, and qualifications', Department of Works and Pensions, Research Report No. 295.

Niederle, M., and Vesterlund, L. (2007). 'Do women shy away from competition? Do men compete too much?' *Quarterly Journal of Economics*, vol. 122, no. 3, pp. 1067–101.

Regenstein, M., Meyer, J. and Dickemper Hicks, J. (1998). 'Job prospects for welfare recipients: employers speak out.' *Urban Institute Brief*, series A, no. A-25.

Schneewind, K.A.,Schröder, G. and Cattell, R.B. (1983). *Der 16-Persönlichkeits-Faktoren-Test (16 PF)*. Bern: Verlag Hans Huber.

Wößmann, L. (2008). 'Entering the teacher workforce in Germany', Report to the Urban Institute in association with its work with the Gates Foundation on the development of human capital in education, University of Munich.

Tables

Table 1

Selection into Variable Pay: Experimental Data

	(1)	(2)
Productivity	-0.002**	-0.002**
	[0.001]	[0.001]
High-School GPA (Abitur grade)	-0.140***	-0.127**
	[0.051]	[0.057]
Female	-0.198***	-0.281***
	[0.069]	[0.080]
Risk attitude	0.055***	0.041**
	[0.019]	[0.021]
Norm-orientation		-0.049
		[0.033]
Ability to work under pressure		0.049*
		[0.027]
Independence		-0.014
		[0.030]
Readiness to take decisions		-0.011
		[0.027]
Extraversion		-0.011
		[0.027]
PR	0.086	0.098
	[0.082]	[0.085]
RS	0.112	0.144*
	[0.081]	[0.086]
Observations	235	218
Pseudo R-squared	0.123	0.136

The table reports marginal effect estimates of Probit models evaluated at the mean of explanatory variables. Standard errors in brackets, * significant at 10%; ** at 5%; *** at 1%. Sample: Participants in Experiment. Dependent variable takes value 1 if subject chooses variable pay. PR is a treatment dummy for piece-rate treatment; RS is a treatment dummy for the revenue-sharing treatment. The productivity measure is the time in seconds that an individual needs to solve a multiplication problem of difficulty level four correctly (see text for details).

Sample: Experimental data.

Table 2

Selection in the Field: SOEP Data

	(1)	(2)	(3)	(4)	(5)	(6)
Average grade	0.004	0.003	0.001	0.019	0.016	0.012
	[0.017]	[0.018]	[0.018]	[0.013]	[0.013]	[0.013]
1 if female	0.162***	0.161***	0.150***	0.134***	0.123***	0.105***
	[0.027]	[0.028]	[0.030]	[0.021]	[0.021]	[0.023]
Age (in years)	0.010***	0.009***	0.008**	0.011***	0.011***	0.011***
	[0.001]	[0.001]	[0.003]	[0.001]	[0.001]	[0.002]
Risk attitude	-0.024***	-0.024***	-0.024***	-0.016***	-0.017***	0.017***
	[0.006]	[0.007]	[0.007]	[0.005]	[0.005]	[0.005]
Trust	0.028**	0.022	0.021	0.050***	0.046***	0.045***
	[0.013]	[0.013]	[0.013]	[0.010]	[0.010]	[0.010]
Positive reciprocity	0.012	0.012	0.012	-0.004	-0.009	-0.009
	[0.016]	[0.017]	[0.017]	[0.011]	[0.012]	[0.012]
Negative reciprocity	-0.013	-0.017	-0.017	-0.021***	-0.022***	0.022***
	[0.010]	[0.011]	[0.011]	[0.008]	[0.008]	[0.008]
Conscientiousness		-0.026*	-0.026*		-0.022**	-0.021*
		[0.016]	[0.016]		[0.011]	[0.011]
Extraversion		-0.020	-0.021		0.008	0.008
		[0.015]	[0.015]		[0.011]	[0.011]
Agreeableness		0.009	0.008		0.014	0.013
		[0.016]	[0.016]		[0.011]	[0.011]
Openness to experience		0.039**	0.040**		0.017	0.018
		[0.016]	[0.016]		[0.011]	[0.011]
Neuroticism		0.006	0.006		0.009	0.009
		[0.015]	[0.015]		[0.010]	[0.010]
Years of full-time experience			0.001			-0.001
			[0.003]			[0.002]
Years of part-time experience			0.004			0.002
			[0.005]			[0.003]
Observations	1521	1502	1502	1521	1502	1502
Pseudo R-squared	0.060	0.064	0.064	0.158	0.161	0.164

The table reports marginal effect estimates of Probit models evaluated at the mean of explanatory variables. Standard errors in brackets, * significant at 10%; ** at 5%; *** at 1%. The dependent variable in Columns (1) - (3) takes value 1 if the respondent works in the public sector and 0 otherwise. The dependent variable in Columns (4) - (5) takes value 1 if the respondent is a teacher (including any type of teaching profession and heads of school) and 0 otherwise. The sample is restricted to all employed individuals younger than 65 who are currently not enrolled in education and who have obtained a schooling degree that qualifies them for attending university. Average grade is average grade in Math and German (at age 17).The measure of positive reciprocity is the individual's average level of agreement to three statements concerning, respectively, willingness to return a favour,

to go out of the way to help somebody who was kind, and undergo personal costs to help someone who was helpful before. The measure of negative reciprocity reflects average agreement to statements concerning willingness to take revenge for a serious wrong, to retaliate for being put in a difficult position, and to respond to an insult with an insult. Answers are always on a scale from 1 to 7, where 1 means "does not apply to me at all" and 7 means "applies to me perfectly." The measure of trust is the standardized principal component of the individual's level of agreement with three different statements regarding the trustworthiness of others (agreement is measured on a four-point scale from "agree completely" to "disagree completely") that were included in the 2003 wave. The measure is obtained from a principal component analysis on the estimation sample.

Sample: SOEP 2004; measures of Big Five and reciprocity based on the 2005 wave; measures of trust based on the 2003 wave.

Table 3

Big Five used in the Socio-Economic Panel

You will probably find that some apply to you perfectly and that some do not apply to you at all. With others, you may be somewhere in between. Please answer according to the following scale: 1 means "does not apply to me at all", 7 means "applies to me perfectly". With values between 1 and 7, you can express where you lie between these two extremes

Factor	Adjectives
Conscientiousness	does a thorough job; tends to be lazy; does things effectively and efficiently
Extraversion	is communicative, talkative; is outgoing, sociable; is reserved
Agreeableness	is sometimes somewhat rude to others; has a forgiving nature; is considerate and kind to others
Openness to experience	is original, comes up with new ideas; values artistic experiences; has an active imagination
Neuroticism	worries a lot; gets nervous easily; is relaxed, handles stress well

Table 4

Personal Characteristics and Type of Teaching Career

	(1) Primary school teacher	(2) Teacher at junior high school	(3) High school teachers	(4) College/ University professors	(5) Other teachers
Average Grade	0.438*	0.445***	-0.279	-0.603**	0.365
	[0.249]	[0.149]	[0.175]	[0.285]	[0.240]
1 if female	1.739***	1.626***	0.141	-0.556	1.427***
	[0.438]	[0.249]	[0.271]	[0.472]	[0.394]
Age (in years)	0.091***	0.102***	0.110***	0.033	0.051***
	[0.022]	[0.013]	[0.015]	[0.021]	[0.019]
Risk attitude	-0.258***	-0.173***	-0.132**	-0.143	0.079
	[0.094]	[0.056]	[0.063]	[0.102]	[0.093]
Trust	0.161	0.486***	0.305**	0.401*	0.273
	[0.202]	[0.125]	[0.135]	[0.217]	[0.182]
Positive reciprocity	-0.195	0.080	0.020	-0.161	-0.451**
	[0.230]	[0.148]	[0.164]	[0.241]	[0.204]
Negative reciprocity	-0.101	-0.263***	-0.093	-0.206	-0.036
	[0.165]	[0.098]	[0.104]	[0.183]	[0.147]
Std. Conscientiousness	-0.428**	-0.094	-0.175	-0.098	-0.104
	[0.201]	[0.133]	[0.144]	[0.234]	[0.205]
Std. Extraversion	0.031	0.087	0.201	-0.402*	0.175
	[0.209]	[0.127]	[0.145]	[0.208]	[0.214]
Std. Agreeableness	0.339	0.126	0.014	0.102	0.074
	[0.230]	[0.134]	[0.150]	[0.251]	[0.210]
Std. Openness to experience	0.080	-0.027	0.072	0.544**	0.480**
	[0.214]	[0.129]	[0.154]	[0.250]	[0.227]
Std. Neuroticism	-0.011	0.165	-0.008	-0.139	0.088
	[0.202]	[0.122]	[0.142]	[0.219]	[0.181]
Constant	-7.493***	-8.279***	-6.765***	-1.902	-5.494***
	[1.948]	[1.218]	[1.349]	[1.862]	[1.715]
Observations			1502		
Pseudo-R-squared			0.158		

Multinomial logit estimates with standard errors in brackets, *** p<0.01, ** p<0.05, * p<0.1. The reference category is non-teaching profession. Teacher at junior high school comprise *Hauptschule* and *Realschule*. High-school teachers comprise *Gymnasium*, *Gesamtschule* and *Berufsschule*. For details on sample restrictions and variables see notes to Table 2.

Sample: SOEP 2004; measures of Big Five based on the 2005 wave.