

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

Armin Falk  
University of Bonn

Thomas Dohmen  
ROA, Maastricht University, IZA, DIW

PEPG 08-06

**Preliminary draft  
Please do not cite without permission**

**Prepared for the CESifo/PEPG joint conference  
“Economic Incentives: Do They Work in Education”  
Insights and Findings from Behavioral Research**

**CESifo Conference Center  
Munich, Germany  
May 16-17, 2008**

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

Armin Falk

University of Bonn, IZA, CESifo and CEPR

Thomas Dohmen

ROA, Maastricht University, IZA, DIW

*Preliminary Version*

## **1. Introduction**

In this paper we analyze the interaction of incentives and worker self-selection, with a particular focus on the education system. Typically economists are interested in how incentives affect behavior for a given workforce where the rationale for providing incentive and performance related pay is to align the interests of principals and agents in order to increase output. However, output does not only depend on incentives but also on the composition of the workforce, i.e., employees' abilities, attitudes and personalities. The point of this paper is that the worker composition is endogenous, due to worker self-selection: Agents with different characteristics feel attracted by different incentives. In this sense providing incentives in firms or organizations has two important effects, an incentive effect per se and a selection effect. These effects can but do not have 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 selected 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 argue whether selection effects, which are expected in response to changing the incentive schemes, will have a positive or negative net effect. What the paper does show, however, is that in as much as individual characteristics of teachers, and employees in general, matter for the success of organizations neglecting this selection is highly problematic.

We proceed in three steps. In the first step we discuss and reanalyze data from a recent laboratory experiment by Dohmen and Falk (2006). The laboratory evidence sets 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, which is either a piece rate, a tournament or a revenue sharing contract. 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 on field data. 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 incentive specific effects difficult. Moreover, productivity measures are often not available or fraught with measurement error and preference parameters and personality indicators are typically not available. An interesting alternative is to study individual firm data, as is done in the seminal paper by Lazear (2000). He reports data from a firm that changed from fixed wages to piece rates, and shows that productivity was not only increased due to an incentive effect but also due to selection effects, which led to a more productive workforce. While the data are very well suited to highlight that productivity sorting is extremely important, the data contain no information about potential sorting effects with respect to preferences or personality.<sup>1</sup>

In the second step we analyze employee selection between the private and the public sector in Germany, using data from the German Socio-Economic Panel

---

<sup>1</sup> Moreover, one cannot rule out the possibility that policy changes in firms are endogenous, which renders causal inferences difficult.

(SOEP). While the public sector is largely characterized by using fixed payment schemes pay for performance schemes are more frequent in the private sector. In addition the unemployment risk, which is basically absent in the public sector but relevant in the private sector, provides strong performance incentives in the latter. 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. In addition we find differences in personality and social preferences. As a measure for personality we use the SOEP version of the so-called Big Five, the most widely used and best established concept to categorize personality along five dimensions: openness to experience, conscientiousness, extraversion, agreeableness and neuroticism.

In the third step we look specifically at selection patterns of teachers. We compare teachers with other employees who have obtained a schooling degree that qualifies them for attending university (Abitur). We run the same regression as before and find that teachers are relatively risk averse compared 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 reciprocity 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 teachers and the composition of teachers in the long-run. Well-being is affected because variable pay exposes teachers to additional risk. Given that teachers, who have selected on the basis of fixed payment schemes are relatively risk averse, job satisfaction is likely to fall. This effect is reinforced if the introduction of variable pay comes along with reduced job security. Moreover, if teachers are genuinely risk averse, introducing incentives may *ceteris paribus* actually not be efficient. The composition is likely to change in the long run simply because differences in remuneration schemes attract 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.

The rest of the paper is organized as follows. Section 2 discusses the laboratory evidence on worker self-selection and sets the stage for our analysis of employees in the private vs. the public sector (Section 3) as well as teachers vs. other employees (Section 4). Section 5 concludes and briefly discusses how changes in incentives to invest in education, e.g. changes in financing of the education system, may affect education decisions of potential students. In particular we discuss our finding that willingness to take risks is positively associated with educational attainment and that individuals with higher discount rates invest less in schooling. These correlations imply that policy interventions that alter uncertainty about future returns to education or that change the timing of costs and benefits affect the level of educational investments and the pool of students. This is particularly relevant as different preferences stem from systematic differences in socio-economic characteristics and parental background.

## **2. Self-selection in the laboratory**

In this section we discuss the experiment by Dohmen und Falk (2006) and reanalyze 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 2006 for a discussion).

In their laboratory experiment Dohmen and Falk (2006) study self-selection into variable pay focusing on the role of individual productivity, risk aversion, social preferences, overconfidence, 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 and a one-digit number with five degrees of difficulty.<sup>2</sup> Difficulty varies because different levels of work memory are involved. This so-called real effort task is useful for the purpose at hand, as it requires no previous knowledge, is simple to explain and measure, provides sufficient heterogeneity and is

---

<sup>2</sup> Examples for the different difficulty levels are: Level 1:  $11 \cdot 9 = ???$ ; Level 2:  $3 \cdot 32 = ???$ ; Level 3:  $6 \cdot 43 = ???$ ; Level 4:  $4 \cdot 68 = ???$ ; Level 5:  $7 \cdot 89 = ???$

a proxy for general cognitive ability. Moreover, learning effects that could possibly confound the results are expected to be absent.

In the first phase of the experiment we measure 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 Abitur grade — 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 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 subjects compete against another subject. 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 Socio-Economic Panel (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 tested in a large-scale field experiment, using a representative sample of 450 German adults (for details see Dohmen et al., 2005). In their paper Dohmen et al. (2005) also show that the survey risk question reliably predicts individuals’ actual risk-taking behavior, as measured by incentive compatible lottery experiments with real money at stake.

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. Table 1 shows how these subjects can be characterized. 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 problem of difficulty level of 4 (Productivity) and the grade at the final high-school exam (Abitur), which ranges from 1.0 (very good) to 4.0 (sufficient). We also control for treatment effects with dummies for the piece-rate treatment (PT) and the revenue sharing treatment (RT), i.e., the tournament treatment is the omitted category. The negative effects of 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.<sup>3</sup> 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.<sup>4</sup>

Column 2 includes gender. 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

---

<sup>3</sup> 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.

<sup>4</sup> For details see Dohmen and Falk (2006).

119 are male. Interestingly women are significantly less likely to sort into variable payment schemes than male subjects. This effect is large; women are about 25 percent less likely to select into variable pay. It suggests that, controlling for productivity, 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 analyze sorting decisions between variable pay schemes that vary the degree of competitiveness, e.g., piece-rate vs. tournament. In these experiments women favor 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: In the piece-rate treatment one can ex ante not be certain about a particular outcome and in the revenue sharing and tournament treatments there is additional uncertainty with respect to the type and output of the other agent. 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. As a measure of risk attitudes we use our simple questionnaire measure. Column 3 shows that subjects who are relatively willing to take risks are significantly more likely to self-select into variable pay, as predicted. As indicated by the marginal effects estimates, being one point more risk averse on the 11-point scale implies a reduced probability of selecting into variable pay of about 5.5 percent. This means that for a given productivity variable pay attracts relatively risk tolerant agents.<sup>5</sup> It is interesting to see that including individual risk attitudes sharply reduces the gender coefficient. 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 SOEP (see Dohmen et al., 2005). Thus at least in part gender specific sorting can be explained with gender specific risk attitudes.

In column 4 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

---

<sup>5</sup> Additional complementary evidence is provided by Bonin et al. (2006), find that individuals who are more willing to take risks are more likely to work in occupations with higher earnings variability.

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 results are based on the outcomes of a personality test that was developed by Hermann Brandstätter and is described in Brandstätter (1988). This so called 16 PA test is a short form of the German-language version of Cattell’s sixteen personality factor questionnaire (16 PF), an internationally well established personality assessment.<sup>6</sup> Column 4 shows that personality sorting is not particularly pronounced in our pooled data set. Only the personality trait “ability to 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. One should note, however, that in our experiment only students took part, and that students are a relatively homogeneous group. Moreover, in Table 1 we have pooled our data: if we look at sorting decisions separately by different incentive schemes and gender, personality differences are more pronounced (see Dohmen and Falk, 2006 for details).

---

<sup>6</sup> The German-language version of the 16 PF was developed by Schneewind, Schröder and Cattell (1983). Its 192 items produce sixteen primary factors of personality, which can in turn be summarized in five independent secondary factors. The short test that was used in the experiment presents subjects with 32 conflictive adjective pairs, which describe traits. For each adjective pair, subjects rate how well an adjective pair describes their character 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.

**Table 1: Selection into variable pay: Experimental data**

	(1)	(2)	(3)	(4)
Productivity	-0.002*** [0.001]	-0.002** [0.001]	-0.002** [0.001]	-0.002** [0.001]
Abitur	-0.099** [0.049]	-0.119** [0.050]	-0.140*** [0.051]	-0.127** [0.057]
Female		-0.255*** [0.065]	-0.198*** [0.069]	-0.281*** [0.080]
Risk attitude			0.055*** [0.019]	0.041** [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	0.094 [0.079]	0.083 [0.080]	0.086 [0.082]	0.098 [0.085]
RS	0.119 [0.078]	0.117 [0.080]	0.112 [0.081]	0.144* [0.086]
Observations	235	235	235	218
Pseudo R-squared	0.05	0.096	0.123	0.136

Standard errors in brackets

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Probit estimates. Marginal effects reported

Sample: Participants in Experiment

Dependent variable takes the 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

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.

### **3. Selection into the private vs. public sector**

Building on our laboratory findings, in this section we move from the lab to the field. We study individual characteristics of employees in the private vs. the public sector before we explicitly study self-selection of teachers.

In Germany, the public sector is characterized by using mostly fixed wages with few elements of variable pay. It is also characterized by relatively low risks concerning income variability and unemployment, but also by relatively low wages compared to the private sector. We would therefore expect a similar sorting pattern as is shown in Table 1. Using data from the Socio Economic Panel (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. We restrict the sample to employees who have completed an Abitur and use final high-school grades in mathematics and German. Risk attitudes are measured by the exact same measure as in Table 1, i.e., we use answers 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 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 using factor analysis (cf. Dohmen et al., 2008a): (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 labeled: *strongly agree, agree somewhat, disagree somewhat, strongly disagree.* Similar to the

survey risk measure the trust measure has been validated in a paid experiment. Fehr et al. (2003) 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 the incentive compatible trust game. 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.<sup>7</sup> Agreement to these statement is indicated on a scale from 1 (meaning "does not apply to me at all") to 7 ("applies to me perfectly").

Results are shown in Table 2, which provides several Probit regressions where the dependent variable is 1 if the respective employee works in the public sector.<sup>8</sup> Column 1 shows a weak and insignificant productivity effect as measured by the average math and German grade at age 17. Controlling for grades, column 2 reveals that significantly more women work in the public sector and that public sector employees are significantly less willing to take risks compared to employees in the private sector. Given the differences with respect to risk and uncertainty in the private vs. the public sector, this is exactly what we would expect. It also corroborates our laboratory findings. We also detect significant differences with respect to social preferences. As shown in column 3, 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 4.

---

<sup>7</sup> 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.*

<sup>8</sup> 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.

**Table 2: Selection into the public sector: SOEP data**

	(1)	(2)	(3)	(4)
Average grade	0.01 [0.015]	0.005 [0.016]	0.003 [0.017]	0.002 [0.018]
1 if female		0.154*** [0.024]	0.160*** [0.027]	0.160*** [0.028]
Age (in years)		0.010*** [0.001]	0.010*** [0.001]	0.010*** [0.001]
Risk attitude		-0.023*** [0.006]	-0.024*** [0.006]	-0.024*** [0.007]
Trust			0.017** [0.008]	0.013 [0.008]
Positive reciprocity			0.013 [0.016]	0.012 [0.017]
Negative reciprocity			-0.013 [0.010]	-0.017 [0.011]
Conscientiousness				-0.026* [0.015]
Extraversion				-0.021 [0.015]
Agreeableness				0.009 [0.016]
Openness to experience				0.040** [0.016]
Neuroticism				0.006 [0.010]
Observations	1784	1780	1517	1498
Pseudo R-squared	0.00	0.053	0.06	0.063

Standard errors in brackets

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Probit model. Marginal effects

The dependent variable takes the value 1 if the respondent works in the public sector and 0 otherwise.

The sample is restricted to all individuals who have obtained a schooling degree that qualifies them for attending university

Average grade is average grade in Math and German (at age 17)

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

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 (facets). 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 criticized. For example it has been pointed out that there is no theory why one should cluster personality reliably into 5 groups (and there are many alternative suggestions using different numbers of factors). Moreover, the Big Five model does not consider important factors of 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: 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

In 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, 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.

#### 4. Selection of teachers

We now turn to the selection profile of teachers. We run the same regressions as in Table 2 but now check 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). Our sample is restricted to individuals who are not older than 65 and are currently neither enrolled in school, vocational training nor university. The sample for analysis covers 1784 individuals among whom 439 are teachers (high school teachers, primary school teachers, university professors and teachers in various other education institutions). The dependent variable in all specifications of Table 4 takes value 1 if the individual has self-selected into a teaching profession

Table 4 shows that the results are similar to those that we obtain when comparing employees from the public with those in the private sector. This should not be surprising as teachers in Germany are typically not paid on a pay for performance basis. Column 1 shows a significant negative productivity effect as measured by the average grade in math and German, which tends to be higher (i.e. worse) for teachers. This effect, however, becomes insignificant once we include other explanatory variables. The marginal effects estimates of the Probit model in column (2) indicate that women are 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. This means that individuals who have chosen to become a teacher under the current payment schemes, which is characterized by fixed salaries and job security, are significantly more risk averse, than those in the comparison group. The coefficient indicates that for each point on the 11-point questionnaire scale the probability changes by 1.3 percent. 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.

**Table 4: Selection of teachers**

	(1)	(2)	(3)	(4)
Average grade	0.027** [0.012]	0.014 [0.012]	0.018 [0.013]	0.015 [0.013]
1 if female		0.133*** [0.019]	0.134*** [0.021]	0.123*** [0.021]
Age (in years)		0.011*** [0.001]	0.011*** [0.001]	0.011*** [0.001]
Risk attitude		-0.013*** [0.004]	-0.016*** [0.005]	-0.017*** [0.005]
Trust			0.030*** [0.006]	0.028*** [0.006]
Positive reciprocity			-0.003 [0.011]	-0.008 [0.012]
Negative reciprocity			-0.021*** [0.008]	-0.022*** [0.008]
Conscientiousness				-0.022** [0.011]
Extraversion				0.008 [0.011]
Agreeableness				0.014 [0.011]
Openness to experience				0.017 [0.011]
Neuroticism				0.009 [0.010]
Observations	1786	1782	1518	1499
Pseudo R-squared	0.003	0.126	0.156	0.160

Standard errors in brackets

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Probit model. Marginal effects

The dependent variable takes the 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 individuals who have obtained a schooling degree that qualifies

them for attending university

Average grade is average grade in Math and German (at age 17)

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

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, being all 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 option, which is 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 among teaching careers might trigger differential sorting patterns, we estimated multinomial logit models. The results, which are shown in Table A1 in the Appendix, largely corroborate our findings from Table 4, 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 latter 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 amount of variable pay across these two sub-groups. In terms of personality primary school teachers are significantly less “conscientious”, which holds true also for university professors, albeit not significantly so. Instead the latter are more “open to experience” and show “less extraversion”. High values in terms of “openness to experience” correspond to “valuing artistic experiences”, “being original and coming up with new ideas” and “having an active imagination”. Low values in “extraversion” imply things like being “reserved” and less “outgoing and sociable” or “communicative and talkative”.

To sum up we find that similar to our laboratory findings systematic self-selection takes place, both between the public and private sectors as well as more specifically when comparing teachers to other occupations. These selection effects are likely to be important when considering the consequences of introducing monetary incentives for employees in the education sector, for at least three reasons. First, changes in incentives 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 characterized 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* relatively risk averse. A well-know insight from incentive theory is the tradeoff between risk preference and the strength of incentives. Take a simple linear incentive scheme

$$w = \alpha * \beta(Z) \quad (1)$$

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

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

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  $Var(X)$  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  $\beta$  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 of Dohmen and Falk (2006) as well as those in Tables 2 and 3 suggests 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 but changing the composition of teachers along this dimension may negatively 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 (Glaeser et al., 1996; Dohmen et al., 2006). Dohmen et al. (2008b) use SOEP data and show that the less people are negatively reciprocal, 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 cannot decide 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.

### **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. This 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 considering the optimality design of incentive schemes and organizations.

Sorting effects are not only relevant for teachers, i.e., the supply side of education but also for students, i.e., the demand side. Changing 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. We conclude this section with a short discussion on some non-trivial selection effects. Note first that the choice of enrolling in school or in higher education is often regarded as an investment in human capital, and understood much like an investment in financial or physical capital. The costs of an investment in human capital accrue in the form of tuition fees, forgone earnings and psychic costs, while the private financial returns of such an investment manifest in a higher stream of future earnings (Becker, 1975; Mincer, 1958 and 1974; Ben-Porath, 1967). Since people trade off current earnings and future earnings when deciding on the optimal amount of education, the human

capital model 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).<sup>9</sup> Using data from the SOEP and the same measures as in Sections 3 and 4, we find a positive and significant correlation between educational attainment and willingness to take risks as well as a negative correlation between educational attainment and smoking, which is often used as a proxy for the discount rate (e.g., Munasinghe and Sicherman, 2006)), and which has been shown to be significantly related to various measures of the discount rates that were derived from monetary and non-monetary intertemporal tradeoffs (Golsteyn, 2007). In other words, risk and time preferences systematically determine investment in education.<sup>10</sup> Therefore changes in the cost structure of education alter the choices of students with different preference endowments differently. The introduction of tuition fees, e.g., will change the intertemporal structure of costs and benefits from educational investment, rendering educational attainment less attractive for relatively impatient students. Even if taxes on future earnings are reduced in a such a way that the introduction of the tuition fee is cost neutral in expected terms, the educational investment becomes more risky when future earnings are uncertain so that relatively risk averse students reduce their investments in education.

More generally, policy interventions have heterogeneous effects on individuals with heterogeneous preferences. It is therefore important to understand to what extent preference heterogeneity is determined by personal and socio-economic

---

<sup>9</sup> 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.

<sup>10</sup> These regressions are available on request. We use data from the 2004 wave of the SOEP, and relate educational investment to measures of risk attitudes and the discount rate. We quantify the amount of educational investments by the highest completed degree which is a more natural measure than years of completed education in the German context (cf. Card, 1999, p. 1806) and create an ordinal variable with six degree categories: no degree, *Hauptschulabschluss*, *Realschulabschluss*, *Fachhochschulreife*, *Abitur*, college or university. The *Hauptschulabschluss* qualifies for enrolling in basic vocational schooling, while the *Realschulabschluss* (or *Fachoberschulreife*) is a more advanced degree that is typically completed after 10 years of schooling and qualifies for entering advanced vocational education and for obtaining additional schooling at university-track high-schools. *Fachabitur* and *Abitur*, are the most advanced pre-university degrees in Germany. These degrees require the completion of an exam, which is taken at the end of university-track high-schools (Gymnasium or Sekundarstufe II of Gesamtschule).

characteristics. 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 even in a world with perfect credit markets and policies that increase uncertainty about returns to education would predominately deter individuals from poorer households from studying. Similarly, a shift from tax financing to tuition based financing of education will particularly affect individuals from poorer backgrounds as such a policy changes the payoff risk.

When we analyze determinants of different risk and time preferences of potential students (age 17 to 20), we do in fact find that preferences depend on the social status of their parents.<sup>11</sup> Our results suggest that children from low-wealth households are not only more risk averse but also have higher discount rates than children from richer families. This offers a potential novel explanation for why disadvantaged children invest less in education, and why social mobility is low. It also suggests that policy changes that render educational investments more risky and less attractive for relatively impatient students will reduce social mobility even further.

---

<sup>11</sup> Regressions are available on request.

## References

- Atkinson, J., and M. Williams (2003). "Employer Perspectives on the Recruitment, Retention and Advancement of Low-Pay, Low-Status Employees." Government Chief Social Researcher's Office, UK.
- Autor, David H. and David Scarborough (2005). "Will Job Testing Harm Minority Workers? Evidence from the Retail Sector." MIT, Department of Economics.
- Becci, Newton, Jennifer Hurstfield, Linda Miller, Rosie Page and Karen Akroyd (2005). "What Employers Look for When Recruiting Unemployed and Inactive: Skills, Characteristics, and Qualifications." Department of Works and Pensions, Research Report No. 295.
- Becker, Gary (1975). *Human Capital*. Columbia University Press, New York.
- Ben-Porath, Yoram (1967). "The production of Human Capital and the Life Cycle Earnings." *Journal of Political Economy*, vol. 75, pp. 352-365.
- Brandstätter, Hermann. (1988): "Sechzehn Persönlichkeits-Adjektivskalen (16 PA) als Forschungsinstrument anstelle des 16 PF. *Zeitschrift für experimentelle und angewandte Psychologie*, 35(3), pp. 370-391.
- Bonin, Holger, Thomas Dohmen, Armin Falk, David Huffman and Uwe Sunde (2007). "Cross-Sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes." *Labour Economics*, vol. 14(6), pp. 926-937.
- Borghans, Lex, Angela Lee Duckworth, James J. Heckman and Bas ter Weel (2008). "The Economics and Psychology of Personality Traits." *Journal of Human Resources*, forthcoming.
- Cadsby, Bram C., Fei Song and Francis Tapon (2007) "Sorting and Incentive Effects of Pay-for-Performance: An Experimental Investigation." *The Academy of Management Journal*, vol. 50(2), pp. 387-405.
- Card, D. (2001). "Estimating the return to schooling: progress on some persistent econometric problems." *Econometrica*, vol. 69, pp. 1127-1160.
- Dohmen, Thomas and Armin Falk (2006). "Performance Pay and Multi-dimensional Sorting: Productivity, Preferences and Gender", IZA DP 2001.
- Dohmen, Thomas, with Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp and Gert G. Wagner (2005). "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally- Validated Survey", IZA DP 1730.
- Dohmen, Thomas, Armin Falk, David Huffman and Uwe Sunde (2006). "The Intergenerational Transmission of Risk and Trust Attitudes", IZA DP 2380.
- Dohmen, Thomas, Armin Falk, David Huffman and Uwe Sunde (2008a). "Representative trust and reciprocity: prevalence and determinants." *Economic Inquiry*, vol. 46(1), pp. 84-90.

Dohmen, Thomas, Armin Falk, David Huffman and Uwe Sunde (2008b). "Homo Reciprocans: Survey Evidence on Behavioral Outcomes", Working Paper, ROA, Maastricht University.

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

Fehr and Falk (2002). "Psychological Foundations of Incentives." *European Economic Review*, vol. 46, pp. 687-724

Fehr, Ernst, Urs Fischbacher, B. v. Rosenblatt, Jürgen Schupp and Gert G. Wagner (2003). "A Nation-Wide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys." Mimeo, Institute for Empirical Research in Economics, University of Zurich.

Gerlitz, Jean-Yves and Jürgen Schupp (2005). "Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP" *DIW Research Notes*, 4.

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

Gould, Eric D., Omer Moav and Bruce Weinberg (2001). "Precautionary Demand for Education, Inequality, and Technological Progress." *Journal of Economic Growth*, vol. 6(4), pp. 285-315.

Hartog, Joop and Luis Diaz-Serrano (2007). "Earnings risk and demand for higher education." *Journal of Applied Economics*, vol. 10(1), pp. 1-28,

Hause, J.C. (1974). "The risk element in occupational and educational choices: comment." *Journal of Political Economy*, vol. 82(4), pp. 803-807.

Lazear, Edward P. (2000). "Performance Pay and Productivity," *American Economic Review*, vol. 90(5), 1346-1362.

Levhari, D. and Yoram Weiss (1974). "The Effect of Risk on the Investment in Human Capital." *American Economic Review*, vol. 64(6), pp. 950-963.

Mincer, Jacob (1958). "Investment in Human Capital and the Personal Income Distribution." *Journal of Political Economy*, vol. 66, pp. 281-302.

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

Munasinghe, Lalith and Nachum Sicherman (2006). "Why Do Dancers Smoke? Smoking, Time Preference, and Wage Dynamics." *Eastern Economic Journal*, vol. 32(4), pp. 595-616.

Niederle, Muriel, and Lise Vesterlund (2007). "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, vol. 122(3), pp. 1067–1101.

Regenstein, Marsha and Jack Meyer and Jennifer Dickemper Hicks (1998). "Job Prospects for Welfare Recipients: Employers Speak Out," *Urban Institute Brief*, Series A (No. A-25).

Schneewind, Klaus A. and Gundo Schröder and Raymond B. Cattell (1983). *Der 16-Persönlichkeits-Faktoren-Test (16 PF)*. Verlag Hans Huber, Bern.

Wößmann, Ludger (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.

## Appendix

Table A1: Personal Characteristics and Type of Teaching Career

		(1)	(2)	(3)	
Primary school teacher	Average Grade	0.593*** [0.228]	0.516** [0.243]	0.432* [0.248]	
	1 if female	1.860*** [0.419]	1.738*** [0.430]	1.741*** [0.438]	
	Age (in years)	0.095*** [0.020]	0.089*** [0.022]	0.092*** [0.022]	
	Risk attitude	- 0.265*** [0.087]	- 0.265*** [0.091]	- 0.258*** [0.094]	
	Trust		0,144 [0.120]	0,1 [0.123]	
	Positive reciprocity		-0,191 [0.209]	-0,192 [0.230]	
	Negative reciprocity		-0,134 [0.154]	-0,1 [0.165]	
	Std. Conscientiousness			-0.427** [0.198]	
	Std. Extraversion			0,032 [0.209]	
	Std. Agreeableness			0,338 [0.230]	
	Std. Openness to experience			0,081 [0.214]	
	Std. Neuroticism			-0,01 [0.202]	
	Constant	- 9.544*** [1.310]	- 7.506*** [1.872]	- 7.605*** [1.956]	
	Teacher at junior high school (Haupt- und Realschule)	Average Grade	0.456*** [0.133]	0.479*** [0.144]	0.439*** [0.147]
		1 if female	1.690*** [0.223]	1.691*** [0.240]	1.624*** [0.249]
		Age (in years)	0.104*** [0.012]	0.102*** [0.013]	0.101*** [0.013]
		Risk attitude	- 0.131*** [0.049]	- 0.172*** [0.053]	- 0.173*** [0.056]
Trust			0.306*** [0.074]	0.294*** [0.076]	

	Positive reciprocity	0,157 [0.134]	0,085 [0.147]	
	Negative reciprocity	- 0.237*** [0.091]	- 0.263*** [0.098]	
	Std. Conscientiousness		-0,099 [0.131]	
	Std. Extraversion		0,087 [0.126]	
	Std. Agreeableness		0,125 [0.134]	
	Std. Openness to experience		-0,025 [0.129]	
	Std. Neuroticism		0,167 [0.122]	
	Constant	- 8.835*** [0.755]	- 9.181*** [1.176]	
	Constant	- 8.461*** [1.223]	- 8.461*** [1.223]	
High school teachers (Gymnasium, Gesamtschule, Berufsschule)	Average Grade	-0,233 [0.157]	-0,275 [0.170]	
	1 if female	0,255 [0.239]	0,136 [0.271]	
	Age (in years)	0.108*** [0.014]	0.113*** [0.015]	
	Risk attitude	-0,086 [0.055]	-0.108* [0.060]	
	Trust		0.192** [0.080]	
	Positive reciprocity		0,023 [0.164]	
	Negative reciprocity		-0,091 [0.104]	
	Std. Conscientiousness		-0,179 [0.142]	
	Std. Extraversion		0,204 [0.145]	
	Std. Agreeableness		0,017 [0.150]	
	Std. Openness to experience		0,076 [0.154]	
	Std. Neuroticism		-0,007 [0.141]	
		Constant	- 7.055*** [0.832]	- 7.210*** [1.288]
		Constant	- 6.931*** [1.356]	- 6.931*** [1.356]

College/University professors	Average Grade	0.703***	-0.623**	-0.604**
		[0.241]	[0.280]	[0.284]
	1 if female	-0,63	-0,681	-0,56
		[0.388]	[0.460]	[0.472]
	Age (in years)	0.035**	0.039*	0,033
		[0.017]	[0.021]	[0.021]
	Risk attitude	-0.141*	-0,148	-0,143
		[0.082]	[0.097]	[0.102]
	Trust		0.288**	0.243*
			[0.128]	[0.132]
	Positive reciprocity		-0,09	-0,159
			[0.228]	[0.241]
	Negative reciprocity		-0,227	-0,206
			[0.166]	[0.183]
	Std. Conscientiousness			-0,095
				[0.231]
	Std. Extraversion			-0.401*
			[0.208]	
Std. Agreeableness			0,103	
			[0.251]	
Std. Openness to experience			0.542**	
			[0.249]	
Std. Neuroticism			-0,137	
			[0.218]	
	Constant	2.732***	-2,349	-2,098
		[1.023]	[1.824]	[1.881]
Other teachers	Average Grade	0,283	0,338	0,351
		[0.210]	[0.232]	[0.237]
	1 if female	1.759***	1.620***	1.430***
		[0.366]	[0.381]	[0.393]
	Age (in years)	0.055***	0.056***	0.051***
		[0.017]	[0.018]	[0.019]
	Risk attitude	0,108	0,128	0,08
		[0.080]	[0.088]	[0.093]
	Trust		0.185*	0,164
			[0.111]	[0.111]
	Positive reciprocity		-0.328*	-0.447**
		[0.185]	[0.204]	
Negative reciprocity		-0,022	-0,034	
		[0.140]	[0.147]	
Std. Conscientiousness			-0,108	
			[0.202]	
Std. Extraversion			0,174	
			[0.214]	

Std. Agreeableness			0,072 [0.209]
Std. Openness to experience			0.477** [0.226]
Std. Neuroticism			0,089 [0.181]
Constant	8.305*** [1.076]	6.673*** [1.637]	5.628*** [1.726]
Observations	1782	1518	1499
Pseudo-R-squared	0,125	0,147	0,157

Standard errors in brackets

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Multinomial logit model

The reference category is non-teaching profession.

The sample is restricted to all individuals who have obtained a schooling degree that qualifies them for attending university

Average grade is average grade in Math and German (at age 17)

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