Juerges, Hendrik; Richter, Wolfram F.; Schneider, Kerstin

Working Paper

Teacher quality and incentives theoretical and empirical effects of standards on teacher quality

CESifo Working Paper, No. 1296

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Juerges, Hendrik; Richter, Wolfram F.; Schneider, Kerstin (2004) : Teacher quality and incentives theoretical and empirical effects of standards on teacher quality, CESifo Working Paper, No. 1296, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/18934

TEACHER QUALITY AND INCENTIVES THEORETICAL AND EMPIRICAL EFFECTS OF STANDARDS ON TEACHER QUALITY

HENDRIK JUERGES
WOLFRAM F. RICHTER
KERSTIN SCHNEIDER

CESIFO WORKING PAPER NO. 1296
CATEGORY 1: PUBLIC FINANCE
OCTOBER 2004

PRESENTED AT CESIFO AREA CONFERENCE ON PUBLIC SECTOR ECONOMICS, MAY 2004

CESifo *Working Paper No. 1296*

TEACHER QUALITY AND INCENTIVES THEORETICAL AND EMPIRICAL EFFECTS OF STANDARDS ON TEACHER QUALITY

Abstract

Applying the theory of yardstick competition to the schooling system, we show that it is optimal to have central tests of student achievement and to engage in benchmarking because it raises the quality of teaching. This is true even if teachers' pay (defined in monetary terms) is not performance related. If teachers value reputation, and if teaching output is measured so that it becomes comparable, teachers will increase their effort. The theory is tested using the German PISA-E data. Our estimates suggest that, despite the flat career profile of German teachers, the quality of teaching tends to be higher in federal states with central exams.

JEL Code: I28.

Keywords: education, teacher quality, central examinations, yardstick competition, matching.

Hendrik Juerges
MEA, University of Mannheim
L13, 17
68131 Mannheim
Germany
juerges@mea.uni-mannheim.de

Wolfram F. Richter
University of Dortmund
LS VWL
44221 Dortmund
Germany
wolfram.richter@wiso.uni-dortmund.de

Kerstin Schneider
University of Wuppertal
Gaußstr. 20
42119 Wuppertal
Germany
kerstin.schneider@wiwi.uni-wuppertal.de

The paper was presented at the CESifo area conference on public sector economics. We are grateful for helpful comments from the participants and in particular to John D. Wilson.

1. Introduction

Teacher quality is viewed as one of the most important inputs in an education production function. Hence, there is a broad consensus that academic achievement of students can be raised if the quality of teachers improves. This insight at hand, politics is challenged to improve the incentives for teachers to perform. A number of countries have changed their schooling institutions or at least conducted large scale experiments to find out how to set the right incentives for teachers.

One way to create incentives for teachers is performance related pay. However, the empirical evidence on the relationship between teacher salaries and teacher quality is surprisingly mixed. For example, Lavy (2002, 2003) finds evidence for positive incentive effects of both performance related salaries and performance related resources given to schools. Moreover, monetary incentives in form of teacher salaries are found to be more cost effective than awarding more resources to the teacher's school. Hanushek, Kain and Rivkin (1999), however, show that salaries and student performance are only weakly related. The composition of teachers within a school district appears to be more affected by characteristics of students than by salary schedules. Apart from tying teacher's pay to the quality of teaching, higher quality could be enforced by stricter certification and licensing provisions. Angrist and Guryan (2003) show that this strategy can fail: the introduction of state-mandated teacher testing in the US has increased teacher wages with no corresponding increase in quality.

Strengthening non-monetary incentives in the schooling system is yet another alternative. This could be simply done by setting common standards, testing students against this standard, and finally making the results public. Teachers will then be motivated to perform well in order to gain non-monetary rewards like reputation or acceptance among colleagues, parents, and students.

In Section 2, we show in a theoretical model that it is optimal to let a teacher's reward (with monetary and non-monetary components) depend on the absolute and relative performance of the teacher's class. To measure performance as an indicator of teacher quality, common standards are needed and students have to be tested against these standards. Measuring the performance of a teacher's class raises the effort put forth by teachers and hence also the academic performance of the students. It is argued that it is efficiency enhancing not only to measure student achievement but to make student test results comparable by controlling for the socio-economic background of the school or the students. Only intelligent benchmarking yields the maximal efficiency gains. Benchmarking in the school system is already practiced, for instance in the US state of California. Public schools

are evaluated based on a so-called Academic Performance Index (API). Each school has to meet a target and is either rewarded for achieving the target or sanctioned for failing to reach the target. Schools are ranked according to their API value. They are ranked within schools of their type, but also, and this picks up the idea of benchmarking, schools are compared to 100 other schools that are similar with respect to demographic characteristics (Similar Schools Rank).

In the present paper, we use data from the German PISA-E study (PISA-extension) to estimate the effects of external standards on teacher quality. Two types of variables are used to measure teacher quality. First, we use subjective measures of teacher and school quality from the student and parents questionnaires. Second, we analyze student performance as measured by the PISA test score to estimate the effect of external standards on achievement.

Estimating the causal effects of central exams is not straightforward, because it is typically decided on the country level whether to have or not to have central exams. Thus within a country there is hardly any variation in exam types which makes it difficult to estimate the effects of central standards using national data[1]. Germany is an exception because, due to its federal structure, there has been a long standing tradition of testing against external standards at the end of secondary schooling in some federal states and of having no standardized tests in others[2]. Hence the German schooling system is suitable to test for the effects of external standards on teacher quality measures and indirect measures like test scores in international tests.

Jürges, Schneider and Büchel (2003) use data from TIMSS-Germany and estimate the effect of central exit exams (CEE) on test scores in Germany with a difference-in-difference estimator. The estimate is positive and significant but smaller than previous studies had suggested. While Jürges et al. (2003) estimate the effect to be at least one third of a school year equivalent using German data only, Wößmann (2002) uses the international TIMSS micro data and estimates the effect to be as much as about one school year equivalent. Here we present a complementary approach to estimate the effect of central standards, focusing on the quality of teachers. We use data from the PISA-E study to show that teachers' performance is in fact better when standards are enforced through central exit exams. In order to identify the causal effects, students in CEE-states and non-CEE states are matched on the

[1]Using international data, the effects could theoretically be estimated (Bishop 1997, 1999; Wößmann 2002) but the drawbacks are manifold (Jürges and Schneider, 2004; Jürges, Schneider and Büchel, 2003).

[2] However, as a result of the unsatisfactory performance of German students in international student achievement studies like PISA or TIMSS, standardized tests will be adopted in almost all of the remaining federal states in the near future.

basis of the propensity score. The results support the predictions from the theoretical model. Teacher quality is higher in states with CEEs.

The paper proceeds as follows: The theoretical argument is developed in Section 2. Section 3 describes the data. In Section 4 we discuss the empirical model and the results, and in Section 5 we briefly summarize the main findings and conclude.

2. The Model

The theoretical literature almost unanimously argues that CEEs and hence central standards improve student performance and might even raise welfare (Costrell, 1997, Effinger and Polborn, 1999). Central exit examinations are purported to function better as incentives for students, teachers and schools than decentralized examinations (e.g. Bishop, 1997, 1999). Students, for example, benefit because the results of CEEs are more valuable as signals on the job market than the results of non-central examinations, simply because the former are better comparable. Furthermore, students who have to meet an external standard at the end of their school career have no incentive to establish a low-achievement cartel in class, possibly with the tacit consent of the teachers. Student test results can be used to monitor teacher and teaching quality on a regular basis. Whether incentives to improve teaching quality, arguably an important factor in the education production function, should come solely from reputation effects on the teacher or school level, or in form of higher pay for better teachers is open to discussion (Hanushek et al., 1999; Lavy, 2002, 2003; Glewwe, Ilias and Kremer, 2003).

The following model describes how teachers determine effort and how a social planner chooses the components of the teacher's reward to maximize a social welfare function. The basic idea is that the planner is interested in setting the right incentives for teachers to put forth effort, which is unobservable. The outcome of teaching, the academic achievement of students, reflects effort to some degree, but achievement is an imperfect measure of effort when classes are not homogenous with respect to their average ability. With heterogeneous classes the planner does not know for sure how much effort the teacher has invested. The literature on yardstick competition shows how a first-best level of welfare can be obtained by competing away the asymmetry of information (Armstrong, Cowan and Vickers, 1994). The following model is an application of yardstick competition to the schooling system.

First consider the teachers decision on teaching effort. Each teacher is allocated to one class *i*. The index *i* thus uniquely identifies teachers and their classes. The average ability of students in class *i* is $\tilde{\theta}_i$ (the tilde denotes stochastic variables). Average ability of the students

differs between classes, but we assume that there is no sorting of students by ability and that the average ability of the class is only known to the teacher but not to the planner. The planner could be the principal of the school, given that schools have autonomy , the community or the government. Let $\tilde{\theta}$ be the benchmark for $\tilde{\theta}_i$. One may think of the average ability of a particular class selected for comparison. Alternatively, $\tilde{\theta}$ could be the average ability of a set of classes against which $\tilde{\theta}_i$ is compared. Since average ability is stochastic and students are not sorted by ability we get $E\tilde{\theta} = E\tilde{\theta}_i = \mu$, $\text{var}\tilde{\theta}_i = \sigma^2$, $\text{cov}(\tilde{\theta}_i, \tilde{\theta}) = r\sigma^2$, and $r>0$. The positive covariance between the ability of students in class *i* and the benchmark ensures that there is no systematic sorting of teachers and students. Thus low ability students cannot always be an excuse for the poor performance of teachers' classes. This relationship is crucial for the argument. Only if the ability of students in a class and its benchmark are positively related, it is meaningful to compare academic achievement and to condition the teacher's reward on the relative academic achievement of the students.

Student achievement, $\tilde{a}_i$ (as measured in e.g. central exams or standardized tests like PISA) depends on the ability of the students and teacher's effort, e_i. In particular we choose an additive structure

$$\tilde{a}_i = \tilde{\theta}_i + e_i .$$

The achievement of the benchmark is denoted by $\tilde{a} = \tilde{\theta} + e$.

The teacher's reward, $\tilde{W}_i$, consists of a basic salary and a bonus that depends on the performance of the own class and also on the performance of the benchmark

$$\tilde{W}_i(a_i, a) = \overline{W} + \alpha\tilde{a}_i - \delta\tilde{a} \text{ , with } \alpha, \delta \geq 0 .$$

Note that the bonus does not have to be a monetary bonus but could be reputation or recognition by students, parents or colleagues. Being in a school with a high reputation can be quite valuable for a teacher. Similarly, being assessed as a (relatively) bad teacher can cause disutility and might set strong incentives to improve by working harder. We choose the interpretation of $\alpha\tilde{a}_i - \delta\tilde{a}$ as non-monetary components of the teacher reward to apply the model to the German schooling system. Teacher's pay in Germany is basically not related to performance but simply rises with the age of the teacher. Thus, the career profile of a German teacher is fairly flat. Nevertheless, some federal states decided to make the quality of teaching visible and hence comparable by testing students centrally, thereby allowing the reputation of a teacher to depend directly on the quality of the output: student achievement.

The parameters α and δ are policy parameters in this model. If they assume strictly positive values, the teacher's reward depends on the absolute and relative performance of her class. If only α is positive, the reward depends on the performance of the own class only, but it is not feasible to compare the performance of the teacher's own class to the performance of the benchmark. Positive values of δ indicate that recognition depends also on the performance of the benchmark. Put differently, if my class performs well, I gain recognition. However, if the benchmark performs well, my results are worth less than if the benchmark performs poorly. If α and δ are both zero, teachers receive a basic, performance independent salary only. This is the case if performance is not measured and no benchmark exists against which to compare the achievement of the teacher or the students, respectively. Benchmarking requires a common standard for measuring achievement, which is enforced by means of central exams.

In the following we show that a social planner would optimally choose positive values for both parameters, α and δ. The choice of some positive α is a direct means to elicit teacher's effort. The choice of a positive value for δ is less obvious and needs to be proven. As we will show, δ is smaller than α in the optimum. However, it is larger the stronger the correlation of the average ability $\tilde{\theta}_i$ and the benchmark $\tilde{\theta}$. Thus benchmarking is socially desirable only to the extent to which comparability of abilities is given.

Teachers derive utility from the expected reward, but utility also depends negatively on the work effort. Reward and effort have to be traded off. Moreover, if teachers are risk averse, they do not like uncertain rewards. We write the teachers expected utility function as

$$E(\tilde{U}_i) = E(\tilde{W}_i - \tfrac{1}{2}e_i^2) - \tfrac{1}{2}\gamma \operatorname{var}(\tilde{W}_i).$$

Using the expressions for $\tilde{W}_i$ and e_i, we get

$$E(\tilde{W}_i - \tfrac{1}{2}e_i^2) = \overline{W} + (\alpha - \delta)\mu + \alpha e_i - \delta e \; - \tfrac{1}{2}e_i^2 \text{ and} \qquad (1)$$

$$\operatorname{var}(\tilde{W}_i) = (\alpha^2 + \delta^2)\sigma^2 - 2\alpha\delta\sigma^2 r.$$

Since the variance does not depend on the effort e_i, teachers determine optimal effort by maximizing (1), which results in $e_i^* = \alpha$. Setting $\tilde{W}_i^* := \tilde{W}_i|_{e_i = e_i^*}$ and assuming symmetry, $e_i^* = e^*$, we obtain

$$E(\tilde{W}_i^* - \tfrac{1}{2}e_i^{*2}) = \overline{W} + (\alpha - \delta)\mu + \tfrac{1}{2}\alpha^2 - \delta\alpha.$$

The social planner decides on the policy parameters, i.e. the structure of the teacher reward. In decentralized systems, the social planner could be the principal of the school, in centralized systems it could be the ministry of education. The social planner maximizes a welfare function of the type

$$G = G(\tilde{a}_i^*, \tilde{W}_i^*)$$

with $\frac{\partial G}{\partial \tilde{a}_i^*} > 0$ and $\frac{\partial G}{\partial \tilde{W}_i^*} < 0$, i.e., the social planner is interested in the academic performance of the students but wants to keep the rewards low. Assuming additivity yields

$$G = E(a_i^* - \tilde{W}_i^*) \quad = \mu + \alpha - \left[\overline{W} + (\alpha - \delta)\mu + \alpha^2 - \delta\alpha\right].$$

The planner maximizes the welfare function by determining the optimal structure of teachers' reward, respecting the participation constraint. Thus she

$$\max_{\overline{W},\alpha,\delta} G \quad s.t. \quad E(\tilde{W}_i^* - \tfrac{1}{2}e^{*2}_i) - \tfrac{1}{2}\gamma \operatorname{var}(\tilde{W}_i^*) = const.$$

The corresponding Lagrangean is

$$\begin{aligned} \Lambda = & [\mu + \alpha - \overline{W} - (\alpha - \delta)\mu - \alpha^2 + \delta\alpha] \\ & + \lambda\left[\overline{W} + (\alpha - \delta)\mu + \tfrac{1}{2}\alpha^2 - \delta\alpha - \gamma\sigma^2(\tfrac{1}{2}\alpha^2 + \tfrac{1}{2}\delta^2 - r\alpha\delta)\right] \end{aligned} \tag{2}$$

Partial differentiation with respect to $\overline{W}$ yields $\lambda = 1$.

Using $\lambda = 1$ in (2) gives

$$\Lambda = (\mu + \alpha) - \tfrac{1}{2}\alpha^2 - \gamma\sigma^2(\tfrac{1}{2}\alpha^2 + \tfrac{1}{2}\delta^2 - r\alpha\delta). \tag{3}$$

Differentiating (3) with respect to δ yields

$$\delta = r\alpha, \tag{4}$$

and finally from the first-order condition with respect to α we get

$$\alpha = \frac{1}{1 + \gamma\sigma^2(1 - r^2)}. \tag{5}$$

Note that $\alpha > 0$, whereas $\delta > 0$ only if $r > 0$. Hence it is always optimal to reward teachers according to the absolute performance of the class. However it is only optimal to reward teachers according to relative academic achievement if comparability can be ensured. The better the comparability as measured by a large value of *r*, the better the benchmark. In case of perfect correlation, *r*=1, the first best, $1 = \delta = \alpha = e_i^*$, is obtained. This raises the issue on how to choose the benchmark against which to compare the achievement of class *i.*

Clearly, if $\tilde{\theta}$ is the average ability of all students in the country, $\tilde{\theta}$ is non-stochastic and r vanishes. As a result $\delta = 0$. The interpretation is that there is little gain in social welfare from benchmarking when comparability cannot be ensured. Vice versa, the more careful the benchmark is chosen with respect to for instance the comparability of the socio-economic background, the more efficient are the results.

If $r<1$, it still pays to reward teachers according to absolute academic achievement, $\alpha > 0$, however the first best is systematically failed. The reason is that teachers are assumed to be risk averse, and the social planner has to account for this as it affects the participation constraint. The more risk averse teachers are, or the larger the variance of students' average ability, the more costly it is to reward teachers according to student achievement.

To summarize the main results of the theoretical model: It is efficiency enhancing to let teachers' reward depend on absolute and relative performance measures based on the academic achievement of students. The requirement for this is a standardized evaluation of student achievement in form of a central exam. Moreover, efficiency gains can be realized if the performance of classes as an indicator of teacher quality is evaluated relative to a good benchmark. This can be achieved by controlling for observables like the socio-economic background of students. In the following empirical part of the paper, we test whether teacher quality is in fact higher with central exams.

3. The Data

The data used in the empirical analysis is drawn from the German PISA-E data[3]. The OECD-Program for International Student Assessment – PISA – aims to assess how much knowledge and skills students approaching the end of compulsory schooling have acquired in the basic fields of reading, mathematics, and science[4]. A total of 32 countries participated in the first assessment in 2000 with the focus of the testing being on the reading literacy of 15 year old students. In 2003 the major domain is mathematical literacy and in 2006 the focus is on scientific literacy. Each country tested between 4,500 and 10,000 students. In Germany, 5,000 students from 219 schools participated in the first PISA test.

[3] The data is available on the website of the German *Kultusministerkonferenz* (Ministries of education of the federal states) under http://www.kmk.org/.

[4] One argument against central tests is the possible tendency of teaching-to-the-test. Clearly, if the teachers know what students need to know when taking the test, they might be tempted to focus exclusively on exam relevant issues. Teaching-to-the-test could in fact be counterproductive. Our data, however, stems from the PISA-study and not from testing achievement based on a national curriculum. Thus teachers in Germany where not familiar with the test content, and, if teaching-to-the-test is a relevant problem at all, it should not affect the PISA results.

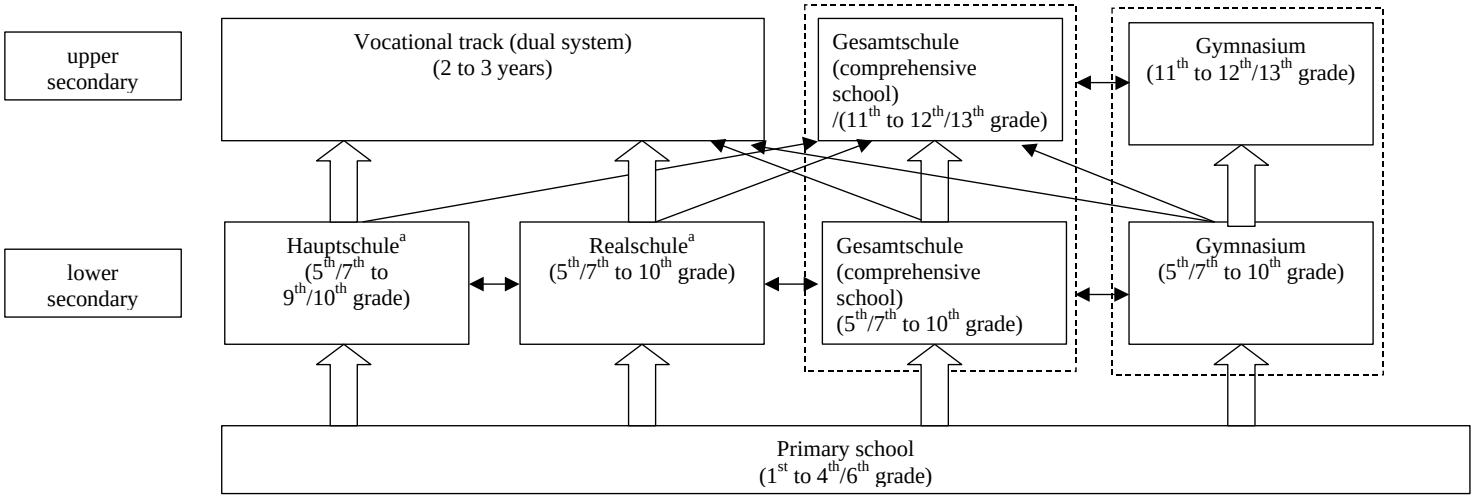
In addition to the international version of PISA, Germany complemented PISA by a national extension, called PISA-E, which was conducted simultaneously with the PISA test. PISA-E is a study of 15 year old students and 9th graders. The international test was supplemented by national test items and the sample size was increased from about 5,000 in the international test to two overlapping samples of 33,809 15 year old students and 33,744 9th graders. The overlap is 47 percent.

Since the information about the state of the student's school is only available in the data of the 9th graders, we are working with that part of the sample to assess the effect of CEEs on the quality of teaching. The published data has information on all 16 states but since the data on students in Berlin and Hamburg are not representative, they are excluded from the analysis.

Before we discuss the practice of CEEs in Germany, we briefly describe the German school system in Figure 1[5]. All children in Germany attend primary school, which covers grades 1 to 4, or in some states grades 1 to 6. There is no formal exit examination at the end of primary schooling. Rather, students are generally allocated to one of the three secondary school types on the basis of the primary school's recommendation. If the primary school's recommendation conflicts with the parents' wishes, however, the final decision about the future course of education lies either with the parents, the secondary school, or the school supervisory authority, depending on the federal state.

The *Hauptschule*, *Realschule* and *Gymnasium* are the three main types of secondary school; each leads to a specific leaving certificate. The *Hauptschule* provides its students with basic general education, and usually comprises grades 5 to 9 (or 10 in some states). The *Realschule* provides a more extensive general education, usually comprising grades 5 to 10. The *Gymnasium* provides an in-depth general education covering both lower and upper secondary level, and usually comprises grades 5 to 13 (or 12 in states in eastern Germany). Depending on their academic performance, students can switch between school types. A fourth type of school is the *Gesamtschule* (comprehensive school). This type of secondary school offers all lower secondary level leaving certificates, as well as providing upper secondary education. It only plays a minor role in most federal states with less than 10 percent of all students attending a comprehensive school.

[5] A detailed description of the German school system can be found in Jonen and Boene (2001).



[a] Some Eastern German states integrate Haupt- and Realschule in a middle school.

Figure 1: A model of the German school system

Table 1: Federal States with CEE by degree

	Lower Secondary					Upper Secondary	
	Hauptschule	*Realschule*	Middle school (Hauptschule + Realschule)	Comprehensive school	Gymnasium	Comprehensive School	High school diploma (Abitur)
Baden-Württemberg (BW)	+	+			+		+
Bavaria (BY)	+	+			–		+
Mecklenburg-W. Pomerania (MV)	–	+		+	–	+	+
Saarland (SA)	–	–		–	–	+	+
Saxony			+		–	+	+
Saxony-Anhalt (ST)			+	+	–		+
Thuringia (TH)			+	+	–	+	+

No CEEs in Berlin, Brandenburg, Bremen, Hamburg, Hesse, Lower Saxony, North Rhine Westphalia, Rhineland-Palatine, and Schleswig-Holstein. Grey cells: school type does not exits; +: CEE; –: no CEE.

As mentioned at the outset, decisions concerning the institutional settings of the schooling systems are largely determined on the level of the federal states in Germany. One prominent example of state-specific institutions is the existence of external standards in form of central exit exams (CEE) that allow to compare the quality of teachers by comparing test results, i.e. the academic achievement of the students.

Central exit examinations are most common at the end of upper-secondary education (see Table 1). In 2000, seven out of the sixteen German federal states had a central *Abitur* (high-school diploma) at the state level. These states are concentrated in the south (Baden-Württemberg, Bavaria, Saarland) and east (Mecklenburg-Western Pomerania, Saxony, Saxony-Anhalt, Thuringia). The other states had decentralized systems, where teachers design problems for exit examinations individually subject to the approval of the school supervisory authority. Six states have central exit examinations at the end of *Realschule* and only four have them at the end of *Hauptschule.*

This unique institutional variation allows to empirically test for the effects of central exit exams on the quality of teachers. However, estimating the effect of CEE is not straightforward for various reasons. Teacher's effort or the quality of teaching is unobservable. PISA-E contains a large set of items that can be used to construct indices of teaching quality, which are not necessarily unrelated. Students evaluated their classes and teachers with respect to several dimensions such as achievement pressure, teacher support, disciplinary climate, clarity of instruction, excessive demands, and teachers' individual orientation. Parents were asked to evaluate teachers' demands and efforts, and their overall satisfaction with the school. In addition to these subjective indicators we also use student test results in PISA-E as a more objective indicator of teacher effort. Unlike in TIMSS, teachers were not interviewed in PISA-E, so that we have no self-assessed measures of teacher effort.

The qualitative teacher variables are listed in Table 2. Here, we only mention the number of items used to construct the indices and their reliability (measured by Cronbach's α). Overall, the reliability of the indices is at acceptable to good levels. A detailed list of all items can be found in the Appendix. Here, we only give a short description:

- *Achievement pressure* measures the frequency with which teachers tell their students to work harder.
- *Teacher support* measures the frequency with which teachers help students when they have problems understanding.

- *Bad disciplinary climate* measures the frequency with which bad discipline among students undermines teaching.
- *Clarity of instruction* measures the frequency with which lessons and exercises are clearly structured.
- *Excess demand* measures the frequency with which students think that teachers ask too much of them.
- *Individual orientation* measures the frequency with which teachers commend below-average students who make progress.

Table 2: Indicators of teacher effort

	Mathematics		German		General	
	# of items	alpha	# of items	alpha	# of items	alpha
Students evaluations of						
achievement pressure	3	.65	3	.58		
teacher support	7	.90	6	.85		
bad disciplinary climate	6	.85	6	.80		
clarity of instruction	5	.65	5	.78		
excess demand	4	.74	4	.73		
individual orientation	3	.77	3	.85		
repetitive exercises	2	.57				
innovative exercises	3	.59				
Parents evaluations of						
school's academic level					1	
teachers' efforts					1	
overall satisfaction with school					1	

Students were asked to evaluate teachers in both mathematics and German classes. For mathematics classes, we have two additional indicators, the frequency of repetitive exercises and the frequency of innovative exercises (i.e. exercises that require to apply skills in changing contexts). Parents' evaluations are measured by answers to single questions on academic level, teachers' effort and overall satisfaction with the school.

Besides the subjective judgements of students and parents we use PISA-E test results as a more objective indicator of teacher effort. Unfortunately, official test scores for individual students are not available in the public use data set. There are official test scores, to be precise, but they have been standardized by federal state, which makes any cross-state comparison impossible. However, we have information on whether a student answered a test item correctly or not for all administered items. It is not possible to reconstruct Rasch scores from this data, but we tried to circumvent this problem by constructing two simple scores[6]. (1) The percentage of correct answers a student has given, (2) the weighted percentage of correct

answers, where the weights reflect the difficulty of each question, and where the difficulty of each item is calculated as one minus the percentage of all students that answered the question correctly.

Let x_{ij} be an indicator variable that is 1 if student *i* answered question *j* correctly, and 0 otherwise. The first score is calculated as

$$S1_i = \frac{\sum_{j \in J_i} x_{ij}}{\sum_{j \in J_i} 1}$$

where J_i denotes the test-book issued to student *i* (there was a total of 9 different test-books per subject with overlapping sets of questions). The second score is calculated as:

$$S2_i = \frac{\sum_{j \in J_i} x_{ij} w_j}{\sum_{j \in J_i} w_j}.$$

with

$$w_j = 1 - \frac{\sum_{i \in I_j} x_{ij}}{\sum_{i \in I_j} 1},$$

where I_j denotes the set of students that answered question *j*.

In order to evaluate the relative performance of our score, we calculated intra-state correlations between the official Rasch scores and our scores. The correlations between the Rasch scores and *S*1 range from .75 to .79. Surprisingly, *S*2 (which takes difference in item difficulty into account) performs a bit worse with correlations between .73 and .78.

Table 3 summarizes raw differences in student and parent-assessed teacher effort and student achievement between states with and without central exams. We have standardized all variables to mean zero and variance one, so that these differences can be interpreted in terms of standard errors. Note also that we report separate results for the sub-sample of lower secondary schools (*Haupt-* and *Realschule*). The students in this sub-sample take central exams at the end of lower secondary schooling, i.e., at the end of grade 9 or 10. We thus

[6] We only use international test items. The German PISA-Extension also has tested students in mathematics and science using a "national" set of items.

expect stronger effects of CEEs in these types of schools than in others, because the central exams are still three to four years in the future.

Table 3: Differences between CEE and non-CEE states, weighted

	Full Sample		Lower Secondary Schools	
Mathematics	mean CEE minus mean non-CEE	t-value difference	mean CEE minus mean non-CEE	t-value difference
achievement pressure	0.077	2.510*	0.058	2.146*
teacher support	-0.028	-0.613	0.014	0.323
bad disciplinary climate	-0.132	-2.476*	-0.144	-3.081**
clarity of instruction	0.051	1.650	0.049	1.992*
excess demand	-0.020	-0.694	-0.032	-1.116
individual orientation	0.020	0.782	0.050	1.568
repetitive exercises	-0.112	-2.568*	-0.121	-2.836**
innovative exercises	0.078	1.801+	0.100	2.738**
mathematics score 1	0.162	1.185	0.147	1.233
mathematics score 2	0.151	1.129	0.132	1.274
German				
achievement pressure	0.045	0.901	0.027	0.746
teacher support	-0.081	-1.573	-0.038	-0.670
bad disciplinary climate	-0.051	-1.123	-0.076	-1.621
clarity	-0.017	-0.500	-0.008	-0.282
excess demand	0.008	0.125	0.014	0.398
individual orientation	-0.004	-0.195	0.013	0.399
reading score 1	0.146	0.882	0.145	0.941
reading score 2	0.135	0.813	0.129	0.901
General				
school's academic level	0.281	4.819**	0.245	4.001**
teachers' efforts	0.018	0.459	0.036	0.769
overall satisfaction with school	0.065	1.761+	0.066	1.262

+ p<0.10; * p<0.05; ** p<0.01, t-values adjusted for clustering.

One major problem in assessing the differences between CEE states and non-CEE states is the calculation of the standard errors. Since the data contains no school or class identifiers, we do not know which students belong to the same primary sampling unit and who are thus evaluating the same teacher. The only information we have is the state, the type of school and the track in dual-track or comprehensive schools. We used this information to create clusters, for which we corrected standard errors. It is very likely that these standard errors are too high (and t-values too low).

The largest differences with respect to mathematics teachers can be found for achievement pressure, disciplinary climate, repetitiveness and innovation, and student test scores. All these differences have the expected sign: teachers in CEE states exert more pressure on their students to perform well, they create a more disciplined climate in class, exercises are less repetitive and have more variety. This results in – inter alii – better test scores in mathematics. Thus students in states with external standards outperform their peers in states without external standards. It turns out, that the differences in test scores are much

smaller than those found in Jürges et al. (2003) for the German TIMSS middle school sample (where the raw difference in mathematics scores was 0.433 standard deviations). Moreover, the raw difference in test scores is most likely not an unbiased estimate of the causal effect of external standards. In Section 4 we discuss this issue in depth and propose an unbiased estimator.

Two more things are worth noting: although lower secondary teachers in CEE states appear to give less support when students have trouble understanding, they seem to take more care of weaker students by appraising their progress more often. Interestingly, teachers are less often perceived as too demanding in CEE states. Contrary to our expectations, we find slightly smaller differences in lower secondary schools.

With the exception of achievement pressure, the differences with respect to German teachers and lessons are qualitatively similar to those for mathematics teachers. Test scores are also higher in CEE states.

The last three rows in Table 3 show differences in parental assessment of their children's schools and teachers. Parents in CEE states tend to think more often that the academic level of the school is too high. The differences in parents' evaluations of teacher effort are only small. This is somewhat at odds with their children’s judgment. Parents are perhaps not well informed about their children's teachers and what happens in their classes. Still, overall satisfaction with their children's school is higher in states that have CEEs.

All differences presented so far are raw differences between schools with and without CEEs. The socio-economic background of the students varies between states and below, we will control for these variations. Table 4 describes these covariates variables by CEE status. A couple of differences are worthwhile mentioning. The first difference to note is that having external standards is very common in the East, a heritage of the former GDR's school system. About 40 percent of all students with CEEs are from East Germany, whereas only 9 percent of those without CEEs are from the East. Another difference between the two groups of federal states is that 15.8 percent of the students in non-CEE states do not speak German at home. The corresponding figure for the CEE states is only 9 percent.

While in both types of states, about the same proportion of students visits the *Gymnasium*, there are large differences with respect to the other school types. Comprehensive schools play virtually no role in CEE states but 15 percent of the students in non-CEE states visit comprehensive schools. The educational background of the students is typically very important for their academic achievement. Here we measure educational background by the parents' formal education, the number of books at home, whether there is classical literature at

home and by the frequency of reading to the child before it was able to read by itself. It turns out that the educational background does not vary systematically between the two kinds of states. To control for the impact of family structures, we include the percentage of children living with single parents, which is about 28 percent in both types of states.

Table 4: Description of covariates (prior to matching), weighted.

	Full Sample		Lower Secondary Schools	
	Mean Non-CEE	Mean CEE	Mean Non-CEE	Mean CEE
Boy	0.488	0.488	0.514	0.512
Age	15.284	15.226	15.412	15.279
Single parent	0.279	0.282	0.270	0.293
Parent's educ:low	0.301	0.283	0.390	0.352
Parent's educ:medium	0.438	0.457	0.467	0.495
Parent's educ:high	0.261	0.260	0.143	0.153
Books: 0-10	0.069	0.056	0.100	0.077
Books: 11-50	0.190	0.188	0.248	0.237
Books: 51-100	0.219	0.230	0.256	0.267
Books: 101-250	0.223	0.239	0.210	0.224
Books: 251-500	0.171	0.169	0.111	0.122
Books: 500+	0.128	0.118	0.075	0.073
Classic Literature	0.428	0.466	0.298	0.353
Read to child: never	0.030	0.020	0.040	0.025
Read to child: rarely	0.078	0.069	0.092	0.088
Read to child: once/month	0.079	0.084	0.094	0.101
Read to child: once/week	0.297	0.329	0.319	0.355
Read to child: daily	0.517	0.498	0.454	0.431
Speaks no German at home	0.158	0.090	0.199	0.111
East	0.091	0.398	0.039	0.397
East*No German	0.002	0.012	0.001	0.015
Hauptschule	0.224	0.244	0.424	0.361
Realschule	0.305	0.433	0.576	0.639
Gymnasium	0.319	0.314		
Gesamtschule	0.151	0.010		
East*Hauptschule	0.006	0.039	0.012	0.057
East*Realschule	0.014	0.230	0.027	0.339
East*Gymnasium	0.028	0.122		
East*Gesamtschule	0.042	0.007		
N	12,416	11,193	6,507	6,946

4. The empirical model and the results

In the following we estimate the effect of external standards on teacher quality. Using German PISA data, the most basic approach to identify the causal effect of CEE on student achievement would seem to estimate *simple differences* between average achievement in CEE states and non-CEE states, controlling for student background and other variables of interest. Simple differences, however, have only limited value because they ignore a potentially confounding effect: the endogeneity of CEEs because of self-selection.

Although it cannot be ruled out completely that parents vote with their feet and move between federal states in order to send their children to schools with or without a central exit examination, this seems to be rather unlikely. We therefore assume that the treatment status is

exogenous given the institutional arrangement in each federal state. However, in the long run institutions can change. The existence of CEEs might reflect unobserved variables such as the electorate's preferences for education, that is parental attitudes towards education and achievement in school. When CEEs are correlated with such attitudes, simple differences between CEE and non-CEE states are a biased measure of the CEE effect.

The attempt to estimate the causal effect of CEE is subject to the fundamental problem of causal inference, namely that it is impossible to observe the individual treatment effect (Holland, 1986). One cannot observe the same student at the same time as being student in a state with and without CEE.

In the present paper, we estimate the causal effect of CEEs using an econometric matching estimator. Matching estimators have recently gained much attention in the labor market literature, in particular in the context of program evaluation (for overviews see e.g. Heckman *et al*. (1998) or Blundell and Costa Dias (2000)). They provide an alternative to instrumental variables when there are no good or convincing instruments. Every attempt to identify causal effects must make use of generally untestable assumptions. In the case of matching estimators the assumption is that the selection into a treatment is completely determined by observable variables and that given the observable variables the selection into the treatment is random (unconfoundedness assumption). Provided that the unconfoundedness assumption holds, we can interpret the assignment of students into CEE and non-CEE states as a randomized experiment (given all observed characteristics), which in turn enables us to identify causal effects of external standards. The simplest form of matching proceeds as follows: For each combination of student characteristics compare the quality of teachers in non-CEE states (the controls). Then compute some average difference with respect to the joint distribution of student characteristics. Of course, the larger the number of variables and the larger the number of possible values, the higher the probability of not having a non-CEE student to compare to a CEE student or vice versa. One solution to this dimensionality problem is to condition the comparison on the propensity score (Rosenbaum and Rubin, 1983), which is just the conditional probability of receiving the treatment given the pre-treatment variables. Rosenbaum and Rubin (1983) show that when the selection into treatment is random given the observables, it is also unconfounded given the propensity score. It is thus possible to compute treatment effects conditional on a one-dimensional index.

Still, when the variables are of high dimensionality, it is often not possible to find members of the treatment group and of the control group with exactly the same propensity score. In order to make propensity score matching feasible, we apply nearest neighbor

matching, i.e. each treated individual is matched with the non-treated individual with the "nearest" propensity score.

Table 5: Covariate Differences in Matched Samples

	Full Sample				Lower Secondary Schools			
	Control	Treated	Diff.	t-value	Control	Treated	Diff.	t-value
Propensity Score	0.665	0.665	0.000	0.000	0.706	0.706	0.000	0.001
Boy	0.488	0.480	-0.008	-0.586	0.514	0.509	-0.005	-0.271
Age	15.215	15.208	-0.007	-0.532	15.266	15.269	0.003	0.174
Single parent	0.290	0.298	0.008	0.669	0.293	0.313	0.021	1.185
Parent's educ:low	0.271	0.281	0.010	0.805	0.357	0.361	0.004	0.219
Parent's educ:medium	0.446	0.447	0.001	0.048	0.482	0.482	0.000	0.008
Parent's educ:high	0.283	0.273	-0.010	-0.866	0.161	0.157	-0.004	-0.287
Books: 0-10	0.057	0.057	0.000	0.054	0.077	0.083	0.006	0.676
Books: 11-50	0.196	0.199	0.003	0.256	0.253	0.255	0.002	0.119
Books: 51-100	0.248	0.234	-0.014	-1.154	0.275	0.271	-0.003	-0.202
Books: 101-250	0.226	0.238	0.012	1.070	0.223	0.217	-0.006	-0.368
Books: 251-500	0.165	0.161	-0.004	-0.410	0.107	0.109	0.002	0.177
Books: 500+	0.108	0.111	0.003	0.347	0.065	0.065	-0.001	-0.058
Classic Literature	0.497	0.495	-0.002	-0.129	0.394	0.380	-0.013	-0.704
Read to child: never	0.015	0.016	0.001	0.303	0.015	0.019	0.004	1.109
Read to child: rarely	0.063	0.065	0.001	0.207	0.084	0.082	-0.002	-0.245
Read to child: once/month	0.089	0.087	-0.002	-0.273	0.106	0.107	0.002	0.137
Read to child: once/week	0.335	0.340	0.005	0.359	0.347	0.368	0.020	1.144
Read to child: daily	0.498	0.493	-0.005	-0.336	0.448	0.424	-0.023	-1.224
Speaks no German at home	0.055	0.051	-0.004	-0.629	0.053	0.062	0.010	1.641
East	0.681	0.688	0.007	0.671	0.722	0.731	0.009	0.697
East*No German	0.019	0.016	-0.003	-0.692	0.012	0.021	0.009	1.891+
Hauptschule	0.150	0.154	0.004	0.425	0.254	0.248	-0.006	-0.454
Realschule	0.463	0.467	0.004	0.310	0.746	0.752	0.006	0.454
Gymnasium	0.365	0.358	-0.007	-0.554				
Gesamtschule	0.023	0.022	-0.001	-0.492				
East*Hauptschule	0.070	0.074	0.004	0.468	0.121	0.119	-0.002	-0.141
East*Realschule	0.376	0.380	0.004	0.252	0.601	0.612	0.011	0.620
East*Gymnasium	0.217	0.215	-0.001	-0.113				
East*Gesamtschule	0.018	0.019	0.001	0.372				
N	4,042	11,193			2,157	6,946		

+ p<0.10; * p<0.05; ** p<0.01

The variables used to calculate the propensity score are the same covariates as described in Table 4. In order to show that the matching procedure has indeed produced a balanced sample of treated (CEE) and control (non-CEE) students, we calculate the means of all covariates in the matched sample and test whether these are different (see Table 5). First, note that the control group in the full sample consists of only 4,042 different non-CEE students. Each of these students contributes on average 11,193/4,042 ≈ 2.77 observations to the control group. The corresponding number in the lower secondary school sample is 3.22. t-values in Table 5 account for this fact. Overall, the matching procedure has been quite successful in creating a balanced sample. The only notable difference between treatment and control group seems to be the proportion of children who do not speak German at home in lower secondary school sample.

The matching estimates are displayed in Table 6. We first comment on the subjective quality measures. It turns out that achievement pressure is perceived to be higher in CEE-

states than in non-CEE states. Only in German classes in lower secondary schools, the difference is insignificant. In the full sample, teacher support is perceived to be worse in CEE states than in non-CEE states. However, when one only looks at the lower secondary schools, one gets the opposite – although insignificant – result. In mathematics classes, the disciplinary climate is clearly better in CEE states, but that does not hold for German classes. The results concerning clarity of instruction are also mixed. Mathematics teachers provide somewhat clearer instructions when students will pass a central exam, but the difference to non-CEE students is not significant. In contrast to mathematics teachers, German teachers in CEE states provide less clear instructions than their colleagues in non-CEE states.

Table 6: Differences between CEE and non-CEE states, nearest neighbor matching estimates

	Full Sample		Lower Secondary Schools	
Mathematics	mean CEE minus mean non-CEE	t-value difference	mean CEE minus mean non-CEE	t-value difference
achievement pressure	0.087	3.009**	0.086	2.107*
teacher support	-0.048	-1.728+	0.048	1.227
bad disciplinary climate	-0.061	-2.276*	-0.093	-2.432*
clarity of instruction	0.031	1.078	0.051	1.304
excess demand	-0.021	-0.837	-0.066	-1.699+
individual orientation	0.036	1.302	0.084	2.137*
repetitive exercises	-0.080	-2.978**	-0.106	-2.570*
innovative exercises	0.006	0.196	0.066	1.565
mathematics score 1	0.076	2.185*	0.191	4.072**
mathematics score 2	0.066	1.910+	0.181	4.078**
German				.
achievement pressure	0.061	2.329*	0.050	1.378
teacher support	-0.051	-1.989+	0.031	0.840
bad disciplinary climate	0.006	0.229	-0.016	-0.399
Clarity	-0.055	-2.138*	-0.034	-0.935
excess demand	-0.014	-0.525	-0.011	-0.312
individual orientation	0.030	1.164	0.011	0.293
reading score 1	0.113	4.313*	0.184	4.939**
reading score 2	0.110	4.270*	0.176	4.974**
Parent's General Assessment				.
school's academic level	0.186	7.249**	0.231	6.351**
Teachers' efforts	-0.100	-3.812**	-0.099	-2.565*
overall satisfaction with school	-0.072	-2.664**	-0.075	-1.898+

+ p<0.10; * p<0.05; ** p<0.01

Demands are perceived as somewhat less excessive in CEE states, but the difference is significant only in lower secondary schools' mathematics classes. Teachers in CEE states are generally more oriented towards individual achievement, that is they show interest in and support the progress of all students, independent of their abilities. The difference is strongest for mathematics teachers in lower secondary schools.

Critics of central exams often claim that students are taught to the test. If that were the case in Germany, one would expect more repetitions, in particular of exercises relevant for the

central exam. However, this seems not to be the case, since mathematics exercises are perceived as less repetitive and more innovative in CEE-states.

Let us finally turn to the parents' view. Parents in CEE states are more likely to say that the academic level of their children's school is too high. At the same time they are less satisfied both with the teachers' effort and the school in general. This result is difficult to interpret. The students' reports of the teachers' behavior suggest that, overall, at least mathematics teachers in CEE-states are better teachers. It is possible that this has only a small effect on the parents' judgement which relates to teachers of all subjects. It might also be possible that parents in CEE states have higher expectations, independent of their observable characteristics that were used to balance our sample.

The indicators of teacher quality discussed so far are subjective measures. A more objective though indirect measure of quality is the level of academic achievement itself, measured by the performance in PISA-E. First note that in the full sample all estimates are positive and significant. Thus, the causal effect of central exit exams is positive. We conclude that students in CEE states perform better because of external standards that are enforced by central exit exams. The qualitative results confirm earlier studies by Jürges et al. (2003) and Wößmann (2002). Second, effects in lower secondary schools are stronger than the average. Given the fact that lower secondary students will pass their exam within a shorter period than the average, this result is consistent with the idea that the effect of central exams is stronger when the exams are in the near future. Third, we consider the size of the estimated CEE effect. Size effects are usually reported in terms of school year equivalents. This is not possible with our sample of 9^{th} graders. Instead, we compare the raw differences in the average scores in Table 3 to the matching estimator in Table 6 to show how much of the raw difference in the performance can be attributed central exit exams In the full sample, the estimated effect of CEEs on the mathematics score is less than half the size of the raw difference in the mean scores, while the matching estimate for the reading score difference is about 80 percent of the raw difference. In the lower secondary school sample, the estimated effect is even larger than the raw difference.

5. Conclusions

The paper has made two contributions to the literature on teacher quality. First we argue that it is optimal to reward teachers depending on the absolute and relative academic achievement of students, because this raises the (unobservable) effort of teachers and efficiency. This is true even if the pay (in monetary terms) is not performance related. If

teachers value reputation, they increase effort if the output of teaching – academic achievement of the students – is measured and published. Consequently, academic achievement of students has to be tested centrally and to be made comparable by using a benchmark. The reward mechanism works best if the benchmark is chosen carefully, controlling for observables like the socio-economic background of the school or the students. Thus, central exams are always expected to yield efficiency gains. If the results are made comparable based on intelligent benchmarks, the positive effects are getting stronger.

Second, we used the German PISA-E data to test whether teacher quality is higher when academic achievement of students is evaluated according to a central standard. One particularity of the German schooling system is its federal structure. Some federal states test standards centrally whereas others do not. Another characteristic is the uniform, performance independent pay of German teachers. Our matching estimates suggest that, despite the flat career profile of German teachers, the quality of teaching tends to be in fact higher in federal states with CEEs. We explain this finding explained by teachers' response to non-monetary rewards like reputation. Only CEE states provide the necessary comparable measurement of the academic achievement of students and hence teachers' effort.

6. References

Armstrong, M., Cowan, S., Vickers, J., 1994, Regulatory Reform, Economic Analysis and British Experience, MIT Press.

Adams, R. /Wu, M., 2002, PISA 2000 Technical Report. OECD: Paris.

Angrist J., J. Guryan, 2003, Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements, NBER Working Paper 9545.

Bishop, J.H., 1997, The Effect of National Standards and Curriculum-Based Exams on Achievement. The American Economic Review 87, 260-264.

Bishop, J.H., 1999, Are national exit examinations important for educational efficiency? Swedish Economic Policy Review 6, 349-401.

Blundell R, Costa Dias M., 2000, Evaluation Methods for Non-Experimental Data. *Fiscal Studies* 21, 427-268.

Costrell, R.M., 1997, Can Educational Standards Raise Welfare?, *Journal of Public Economics* 65, 271-293.

Effinger M.R., Polborn, M.K., 1999, A Model of Vertically Differentiated Education, *Journal of Economics* 69, 53-69.

Glewwe, P., Ilias, N., Kremer, M., 2003, Teacher Incentives, NBER Working Paper 9671.

Hanushek, E.A., Kain, J. F. and Rivkin, S. G., 1999, Do Higher Salaries Buy Better Teachers? NBER Working Paper 7082.

Holland PW, 1986, Statistics and Causal Inference. *J Am Stat Assoc* 81; 945-960.

Heckman J.J., Ichimura H., Todd P., 1998, Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Rev Econ Stud* 64: 605-654.

Jonen, G., Boele, K., 2001, *The Education System in the Federal Republic of Germany 2000.* German EURYDICE Unit, Bonn. http://www.kmk.org/dossier/dossier_2000_engl_e-book.pdf [2002, July 1].

Jürges, H., Schneider, K., 2004, International Differences in Student Achievement: An Economic Perspective, *German Economic Review* 5, 357-380.

Jürges, H., Schneider, K., Büchel, F., 2003, The Effect of Central Exit Examinations on Student Achievement: Quasi-Experimental Evidence from TIMSS Germany, CES Working Paper 939.

Lavy, V., 2002, Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement, *Journal of Political Economy* 110, 1286-1317.

Lavy, V. 2003, Paying for Performance: The Effect of Individual Financial Incentives on Teachers' Productivity and Students' Scholastic Outcomes, Working Paper, The Hebrew University of Jerusalem and CEPR.

Rosenbaum, P.R. and D.B. Rubin, 1983, The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70: 41-55.

Wößmann, L., 2002, *Central Exams Improve Educational Performance*: International Evidence. Kiel Discussion Papers 397.

Table A1: Items used to generate teacher effort indices

Student variables	
Achievement pressure	The teacher wants students to work hard. The teacher tells students that they can do better. The teacher does not like it when students deliver sloppy work.
Teacher support	The teacher shows an interest in every student's learning. The teacher gives students an opportunity to express opinions. The teacher helps students with their work. The teacher continues teaching until the students understand. The teacher does a lot to help students. The teacher helps students with their learning. [Mathematics only:] The teacher gives helpful advice for my work
Bad disciplinary climate	The teacher has to wait a long time for students to quiet down. Students cannot work well. Students don't listen to what the teacher says. Students don't start working for a long time after the lesson begins. There is noise and disorder. At the start of class, more than five minutes are spent doing nothing.
Clarity	The teacher gives clear instructions what to do Everything that we do is well planned There are specific rules that we must adhere to The teacher tells us at the beginning of the lesson what to do The teacher summarizes what was done in the previous lesson
Excess demand	Time is too short to finish my work The things we do are too difficult for me The teacher tells us things that I do not understand You stop listening because you do not understand anything
Individual orientation	Our teacher acknowledges improvements even if students are below average When I really make an effort, the teacher commends me even if others are still better than me Our teacher also commends weak students when they make improvements.
Repetitive exercises	We make little progress because we repeat so much We always do the same exercises
Innovative exercises	By some of our exercises, you can really see if you have understood the topic We often apply what we learn to new topics You have to pay close attention because the exercises are similar but always a bit different
Parent variables	
School's academic level	How do you rate the academic level of your child's school – far too low, too low, about right, too high or far too high?
Teachers' efforts	How much do teachers exert themselves for their students – not at all, a little bit, somewhat, much, or very much?
Overall satisfaction with school	How satisfied are you with your child's school – very dissatisfied, dissatisfied, neither dissatisfied nor satisfied, satisfied, or very satisfied?

CESifo Working Paper Series

(for full list see www.cesifo.de)

1233 Cheng Hsiao and M. Hashem Pesaran, Random Coefficient Panel Data Models, July 2004

1234 Frederick van der Ploeg, The Welfare State, Redistribution and the Economy, Reciprocal Altruism, Consumer Rivalry and Second Best, July 2004

1235 Thomas Fuchs and Ludger Woessmann, What Accounts for International Differences in Student Performance? A Re-Examination Using PISA Data, July 2004

1236 Pascalis Raimondos-Møller and Alan D. Woodland, Measuring Tax Efficiency: A Tax Optimality Index, July 2004

1237 M. Hashem Pesaran, Davide Pettenuzzo, and Allan Timmermann, Forecasting Time Series Subject to Multiple Structural Breaks, July 2004

1238 Panu Poutvaara and Andreas Wagener, The Invisible Hand Plays Dice: Eventualities in Religious Markets, July 2004

1239 Eckhard Janeba, Moral Federalism, July 2004

1240 Robert S. Chirinko, Steven M. Fazzari, and Andrew P. Meyer, That Elusive Elasticity: A Long-Panel Approach to Estimating the Capital-Labor Substitution Elasticity, July 2004

1241 Hans Jarle Kind, Karen Helene Midelfart, Guttorm Schjelderup, Corporate Tax Systems, Multinational Enterprises, and Economic Integration, July 2004

1242 Vankatesh Bala and Ngo Van Long, International Trade and Cultural Diversity: A Model of Preference Selection, July 2004

1243 Wolfgang Eggert and Alfons J. Weichenrieder, On the Economics of Bottle Deposits, July 2004

1244 Sören Blomquist and Vidar Christiansen, Taxation and Heterogeneous Preferences, July 2004

1245 Rafael Lalive and Alois Stutzer, Approval of Equal Rights and Gender Differences in Well-Being, July 2004

1246 Paolo M. Panteghini, Wide vs. Narrow Tax Bases under Optimal Investment Timing, July 2004

1247 Marika Karanassou, Hector Sala, and Dennis J. Snower, Unemployment in the European Union: Institutions, Prices, and Growth, July 2004

1248 Engin Dalgic and Ngo Van Long, Corrupt Local Government as Resource Farmers: The Helping Hand and the Grabbing Hand, July 2004

1249 Francesco Giavazzi and Guido Tabellini, Economic and Political Liberalizations, July 2004

1250 Yin-Wong Cheung and Jude Yuen, An Output Perspective on a Northeast Asia Currency Union, August 2004

1251 Ralf Elsas, Frank Heinemann, and Marcel Tyrell, Multiple but Asymmetric Bank Financing: The Case of Relationship Lending, August 2004

1252 Steinar Holden, Wage Formation under Low Inflation, August 2004

1253 Ngo Van Long and Gerhard Sorger, Insecure Property Rights and Growth: The Roles of Appropriation Costs, Wealth Effects, and Heterogeneity, August 2004

1254 Klaus Wälde and Pia Weiß, International Competition, Slim Firms and Wage Inequality, August 2004

1255 Jeremy S. S. Edwards and Alfons J. Weichenrieder, How Weak is the Weakest-Link Principle? On the Measurement of Firm Owners' Control Rights, August 2004

1256 Guido Tabellini, The Role of the State in Economic Development, August 2004

1257 François Larmande and Jean-Pierre Ponssard, EVA and the Controllability-congruence Trade-off: An Empirical Investigation, August 2004

1258 Vesa Kanniainen and Jenni Pääkkönen, Anonymous Money, Moral Sentiments and Welfare, August 2004

1259 Panu Poutvaara and Andreas Wagener, Why is the Public Sector More Labor-Intensive? A Distortionary Tax Argument, August 2004

1260 Lars P. Feld and Stefan Voigt, Making Judges Independent – Some Proposals Regarding the Judiciary, August 2004

1261 Joop Hartog, Hans van Ophem, and Simona Maria Bajdechi, How Risky is Investment in Human Capital?, August 2004

1262 Thomas Eichner and Rüdiger Pethig, Efficient Nonanthropocentric Nature Protection, August 2004

1263 David-Jan Jansen and Jakob de Haan, Look Who's Talking: ECB Communication during the First Years of EMU, August 2004

1264 David F. Bradford, The X Tax in the World Economy, August 2004

1265 Hans-Werner Sinn, Migration, Social Standards and Replacement Incomes. How to Protect Low-income Workers in the Industrialized Countries against the Forces of Globalization and Market Integration, August 2004

1266 Wolfgang Leininger, Fending off one Means Fending off all: Evolutionary Stability in Submodular Games, August 2004

1267 Antoine Bommier and Bertrand Villeneuve, Risk Aversion and the Value of Risk to Life, September 2004

1268 Harrie A. A. Verbon and Lex Meijdam, Too Many Migrants, Too Few Services: A Model of Decision-making on Immigration and Integration with Cultural Distance, September 2004

1269 Thomas Eichner and Rüdiger Pethig, Economic Land Use, Ecosystem Services and Microfounded Species Dynamics, September 2004

1270 Federico Revelli, Performance Rating and Yardstick Competition in Social Service Provision, September 2004

1271 Gerhard O. Orosel and Klaus G. Zauner, Vertical Product Differentiation When Quality is Unobservable to Buyers, September 2004

1272 Christoph Böhringer, Stefan Boeters, and Michael Feil, Taxation and Unemployment: An Applied General Equilibrium Approach, September 2004

1273 Assaf Razin and Efraim Sadka, Welfare Migration: Is the Net Fiscal Burden a Good Measure of its Economics Impact on the Welfare of the Native-Born Population?, September 2004

1274 Tomer Blumkin and Volker Grossmann, Ideological Polarization, Sticky Information, and Policy Reforms, September 2004

1275 Katherine Baicker and Nora Gordon, The Effect of Mandated State Education Spending on Total Local Resources, September 2004

1276 Gabriel J. Felbermayr and Wilhelm Kohler, Exploring the Intensive and Extensive Margins of World Trade, September 2004

1277 John Burbidge, Katherine Cuff and John Leach, Capital Tax Competition with Heterogeneous Firms and Agglomeration Effects, September 2004

1278 Joern-Steffen Pischke, Labor Market Institutions, Wages and Investment, September 2004

1279 Josef Falkinger and Volker Grossmann, Institutions and Development: The Interaction between Trade Regime and Political System, September 2004

1280 Paolo Surico, Inflation Targeting and Nonlinear Policy Rules: The Case of Asymmetric Preferences, September 2004

1281 Ayal Kimhi, Growth, Inequality and Labor Markets in LDCs: A Survey, September 2004

1282 Robert Dur and Amihai Glazer, Optimal Incentive Contracts for a Worker who Envies his Boss, September 2004

1283 Klaus Abberger, Nonparametric Regression and the Detection of Turning Points in the Ifo Business Climate, September 2004

1284 Werner Güth and Rupert Sausgruber, Tax Morale and Optimal Taxation, September 2004

1285 Luis H. R. Alvarez and Erkki Koskela, Does Risk Aversion Accelerate Optimal Forest Rotation under Uncertainty?, September 2004

1286 Giorgio Brunello and Maria De Paola, Market Failures and the Under-Provision of Training, September 2004

1287 Sanjeev Goyal, Marco van der Leij and José Luis Moraga-González, Economics: An Emerging Small World?, September 2004

1288 Sandro Maffei, Nikolai Raabe and Heinrich W. Ursprung, Political Repression and Child Labor: Theory and Empirical Evidence, September 2004

1289 Georg Götz and Klaus Gugler, Market Concentration and Product Variety under Spatial Competition: Evidence from Retail Gasoline, September 2004

1290 Jonathan Temple and Ludger Wößmann, Dualism and Cross-Country Growth Regressions, September 2004

1291 Ravi Kanbur, Jukka Pirttilä and Matti Tuomala, Non-Welfarist Optimal Taxation and Behavioral Public Economics, October 2004

1292 Maarten C. W. Janssen, José Luis Moraga-González and Matthijs R. Wildenbeest, Consumer Search and Oligopolistic Pricing: An Empirical Investigation, October 2004

1293 Kira Börner and Christa Hainz, The Political Economy of Corruption and the Role of Financial Institutions, October 2004

1294 Christoph A. Schaltegger and Lars P. Feld, Do Large Cabinets Favor Large Governments? Evidence from Swiss Sub-Federal Jurisdictions, October 2004

1295 Marc-Andreas Mündler, The Existence of Informationally Efficient Markets When Individuals Are Rational, October 2004

1296 Hendrik Jürges, Wolfram F. Richter and Kerstin Schneider, Teacher Quality and Incentives: Theoretical and Empirical Effects of Standards on Teacher Quality, October 2004