I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12200

Selecting or Rewarding Teachers? International Evidence from Primary Schools

Michela Braga
Daniele Checchi
Christelle Garrouste
Francesco Scervini

MARCH 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12200

Selecting or Rewarding Teachers? International Evidence from Primary Schools

Michela Braga
Bocconi University

Daniele Checchi
University of Milan and IZA

Christelle Garrouste
Paris 12 Val de Marne University

Francesco Scervini
University of Pavia

MARCH 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Selecting or Rewarding Teachers? International Evidence from Primary Schools*

Using data from three waves of PIRLS, this paper examines the effect of teacher quality on fourth-grade students' literacy test scores by exploiting variations induced by reforms in teachers' selection and/or reward schemes. We construct an original data set of relevant reforms taking place at the national level over the last century and affecting the working conditions of primary school teachers, matching them by the year they entered the profession. After showing that teacher experience/age and qualification are significantly correlated with student competencies, we study the correlation between teacher working conditions (including recruitment, pay and retirement policies) and pupil achievement. Our identifying assumption is that the impact of reforms dissipates with the distance between the reform's introduction and entry into the profession. The results point to a more selective recruitment process and, to a lesser extent, more generous reward policies as effective ways to enhance student performance.

JEL Classification: H52, I21, I28, J44

Keywords: student achievements, PIRLS, teacher recruitment process, teacher pay

Corresponding author:
Daniele Checchi
University of Milan
Via Festa del Perdono 7
Milan, 20122
Italy

E-mail: daniele.checchi@unimi.it

* Paper presented at the II SISEC conference (Milan, January 2018), EPCS annual meeting (Rome, April 2018), XXX SIEP annual meeting (Padua, September 2018), IX Workshop on Economics of Education (Barcelona, September 2018) and Fondazione Agnelli (Turin, December 2018). We thank all the participants for their helpful comments. All errors are our own responsibility.

1. Introduction

In terms of the years of schooling and student achievements, increasing the quantity and quality of education is a central goal of policy makers, and finding effective tools to achieve this goal is a current hot issue in academic debates. However, nowadays, especially in advanced economies, more so than increasing access to education, it has become even more important to develop the right skills and to give knowledge to students. Poor education quality penalises students for life because it translates into worse life prospects in terms of lower lifetime earnings, more trouble adapting to modern knowledge-based economies and higher unemployment probability. Because educational failure is costly for both the individual and society, enhancing educational standards is a priority in the policy agenda of many countries.

It is not simple to identify which factors contribute to shaping a good schooling environment. The quality of the educational systems does not depend on a single factor but rather on a variety of elements, among which curriculum, the learning environment, organisation of class activities and teaching techniques are some of the key ones. Teachers with different competencies can clearly influence these elements and hence represent a key input for the educational process. The quality of the teachers is an essential element in determining schools' quality. Based on administrative data, a growing body of empirical literature shows that teachers' quality matters and is the most important school input for predicting students' and adult students' learning gains (Hanushek, 2011; Chetty et al., 2014a, 2014b). Nevertheless, there is still an open debate about what defines teacher quality and how to attract or retain high-quality teachers (Pelayo and Brewer, 2010).

Teacher quality is clearly a combination of observable and unobservable characteristics. Empirical studies, though, can only focus on the observable characteristics (i.e., credentials, experience accumulated on the job, formal and informal training, etc,), which tend to be weakly correlated with teachers' contributions to students' achievements. As a consequence, the unobservable features, such as ability or effort, emerge as significantly correlated with teacher quality.[1]

Because the costs of teaching staff represent, on average, two-thirds of the total schooling expenses and more than 80% of staff compensation in OECD countries (OECD, 2018), from a policy perspective, it is essential to identify the most effective way to recruit and motivate the best teachers. Two possible drivers for attaining this goal are represented by the selection or remuneration practices that are implemented by school principals. However, because both selection and pay policies could be correlated with unobservable characteristics, evaluating which one is the most effective tool is a difficult task. Indeed, the presence or absence of monetary incentives in the profession can induce positive or negative self-selection of individuals. On top of this, it is not clear what the relationship is between the two policy instruments. Namely, it is hard to appraise whether selection and

[1] Goldhaber et al. (1999) estimate that the share of teacher effects because of unobservable characteristics can be as high as 97%.

remuneration are substitutive or complementary tools for enhancing teachers' quality. On the one hand, selectivity should attract better-endowed teachers, both in terms of observable and unobservable traits. On the other hand, merit pay wage policies should attract people who expect to benefit the most from this scheme, but it is hard to anticipate whether they are more able or just greedier than average. As a result, it is almost impossible to predict the overall effect on students' achievement because 'selection' and 'incentive' effects may work in opposite directions.

The current paper contributes to the debate on the quality of educational systems in varying ways. First, different from most previous literature that focuses on secondary schools, we consider the role exerted by teachers on the formation of cognitive skills in primary schools, the first and most essential stage in competence acquisition. Second, a newly assembled data set on the institutional features relevant for primary school teachers over the last century allows for exploiting the exogenous variations in these features to identify in a causal sense the impact of teacher quality on pupil performance. Third, as far as we know, this is the first paper that explores the issue of substitutability and complementarity among alternative policy measures meant to enhance school quality. In addition, although most scholars focus on a single country and on a single reform, we conduct our analysis from a comparative perspective, and we exploit the results from an international comparable standardised assessment that combines information on students' performance with information on their teachers. Obviously, the chosen approach has a trade-off between the number of considered dimensions and their level of detail.

Namely, we use three waves of the Progress on International Reading and Literacy Surveys (PIRLS) conducted over the last decade. We combine these data with an original data set that includes information on the reforming process affecting primary school teachers along four main dimensions of the profession (recruitment, working conditions, pay and retirement) over the last century for around 50 countries; this is combined with information on the teaching framework for primary teachers in the year of the students' assessment. Then, we explore the relationship between teacher working conditions and pupil achievement. More specifically, we proxy individual teachers' quality with group averages, where the groups are identified according to different reforms that have modified the selection and/or the reward of teachers; these measures are then correlated to test the scores of fourth graders. Our identification strategy relies on temporal and geographical variations in the institutional arrangements, here controlling for individual, class, school and teacher characteristics. This strategy is not new, but so far, it has focused only on one type of reform at a time (Braga et al., 2013); however, we take into account the fact that a single cohort of students may be affected by teachers hit by different types of reforms, which can be identified according to teacher's age.

The econometric analysis indicates that policies can significantly improve students' performance via a possible change in teacher quality because of the introduction of a more selective recruitment process and/or more generous pay conditions. Our results have policy implications both in the short and long run. In the short run, being able to attract

and retain the best teachers is a cost-effective strategy. However, it can also have positive spillovers in the long run. Indeed, basic and essential reading, writing and numerical abilities are formed and developed in primary education. These three abilities are fundamental for one's intellectual capabilities in future life. Without them, any investment in skill formation in the subsequent stages of a student's educational process would be costlier and less effective. Indeed, any significant human capital accumulation requires solid foundations built in the early stages of the learning process.

The rest of the current paper is organised as follows: Section 2 reviews the relevant literature. In Section 3, we present the data used for our analysis, while Section 4 is devoted to the empirical strategy. The results are discussed in section 5, while in Section 6, we perform robustness checks and sensitivity analysis to corroborate our results. Finally, Section 7 concludes the paper.

2. Literature Review: The State of the Art

The current paper is linked to two main strands of literature: the one analysing practices to enhance the quality of educational systems and the other on the link between teachers' features and students' performance.

The literature on Human Resource management indicates that there are three main goals for any employer regarding manpower: recruiting (which implies attracting and selecting), motivating and then retaining the employees to achieve their highest level of productivity (Lazear and Gibbs, 2007). Because the ability to teach is a job-specific human capital, retaining teachers is usually not a main concern for school principals and policy makers at large although in the literature, alternative opinions are considered (Moor Johnson, 2006). On the contrary, teacher selection and motivation are at the core of any attempt to improve the quality of educational systems. For these reasons, before moving on to the empirical analysis, we review the existing literature about these two dimensions.

One of the main concerns for policy makers is the selection of teachers who have the appropriate skills to teach. The first step to increase teachers' quality can be obtained at the time of their selection into teacher colleges. However, there is evidence that teacher recruitment is influenced by outside options created by business cycles. For instance, Bacolod (2007) shows that the U.S. experienced a marked decline in the quality of young women entering teaching between 1960 and 1990, contrasting with a simultaneous increase in the quality of those who became professionals. Nagler, Piopiunik and West (2015) obtain analogous results for more recent years by exploiting business cycle conditions at a teacher's start of his or her career as a source of exogenous variation in the outside options of potential teachers. Similarly, Falch et al. (2009) measure teacher shortages in Norway as the share of teachers without certified credentials, finding a negative relationship between teacher shortages and regional unemployment rates in the period from 1981–2002; they explain this effect as a cause of the centralised and rigid pay system in the public sector that tends to reduce labour supply and lead to shortages of qualified personnel. Hence,

according to these authors, the teaching profession would remain a residual one because of the lack of career advancement, leading to a counter-cyclical selection into teacher training: namely, the more favourable the economy is, the lower the probability will be for the best students to opt for a teaching career (see also Dolton et al., 2003). An alternative explanation could rely on the rigidity of salaries for teachers compared with other highly qualified professionals. There is evidence that teacher wage premium and wage dispersion have little effect on the quality mix of teaching applicants. For instance, in a Venezuelan context, Ortega (2010) shows that most students' preferences for teaching are unresponsive to wage levels relative to other occupations and to wage growth prospects within teaching. If wages have a positive effect on the performance of educational systems, it is unlikely to go through the selection of the most talented individuals. Despite this, a study by Dolton and Marcenaro-Gutierrez (2011) on teachers' pay differentials across 39 OECD countries reveals that recruiting more talented individuals into teaching and permitting quicker salary advancements have a positive effect on pupil outcomes.

Thus, given the difficulty of ensuring that the best candidates enter the teaching profession, it becomes even more crucial to ensure the highest quality in teacher training programmes. Unfortunately, the literature does not report any consistent relationship between the level of credentials of teachers and corresponding student achievement (Hanushek and Rivkin, 2006). For instance, on the one hand, the study by Santibañez (2006) on student achievement in Mexico finds a small positive relationship between teacher test scores and average student achievement scores. In a study on the effect of teacher certification on Swedish students' grade point average (GPA), Andersson et al. (2011) show that a one percentage point increase in the share of noncertified teachers is expected to decrease students' GPA by 1.8 standard deviations per year. On the other hand, Harris and Sass (2011) analyse the effects of various types of education and training on the productivity of teachers in promoting student achievement in U.S. schools: they do not find any evidence that teacher preservice (undergraduate) training or college entrance exam scores are related to their productivity. Moreover, the study by Kane et al. (2008) on the effectiveness of recently hired teachers in the New York City public school system shows that on average, the initial certification status of a teacher has only small impacts on student test performance. Along the same line, looking at teacher qualifications and student achievement in Los Angeles primary schools, Buddin and Zamarro (2009) reveal that neither the teacher licensure[2] test scores nor the possession of an advanced degree are related to student achievement. Although selecting skilled people into teaching is a key step towards more effective schools, the academic research indicates that it can be improved by having appropriate reward policies acting as a motivating device. In fact, a common concern of policy makers is the retention of the best teachers and keeping them teaching. Indeed, the cost of a high turnover of high-quality teachers goes beyond a loss in student performance. As reported by Watlington et al. (2010), when high-quality teachers leave the

[2] '*Teacher licensure is a regime where schools are forbidden from hiring teachers who have not completed a program of study in a teacher education program and/or other preparation requirements*' (Arias and Scafidi, 2009).

classroom, there is a significant negative effect on both the students' and the school's performance.

If good teachers are to be retained in the teaching profession and supported in doing their work – and doing it well – they should have a workplace that promotes their efforts in a variety of ways (Moor Johnson, 2006). Since the 1980s, the United States and UK have passed measures to implement performance-based incentives, that is, monetary benefits to teachers and/or school principals, who are considered the best according to the level of (or the variation in) their student achievements (Holanda et al., 2008; Hanushek and Rivkin, 2006; Schächter and Thum, 2004). However, these policy measures have proven to have contradictory effects. Although Atkinson et al. (2009) find that a performance-related pay scheme implemented in the UK did improve test scores and the value added increased on average by about 40% of grade per pupil, Ballou (2001) shows that, in the case of the United States, efforts to implement merit pay in public education have generally been unsuccessful, mainly because of the opposition from teachers and teachers' unions. In Israel, Lavy (2015) reports persistent gains in labour market achievements of students whose teachers have been exposed to pay-to-performance schemes. However, because performance-based incentives are not easy to introduce in public schools, most countries have instead opted for reforms that unconditionally increase the level of teacher salaries. Overall, these measures have been found to be significantly correlated to student achievement (e.g., Dolton and Marcenaro-Gutierrez, 2011; Boarini and Ludemann, 2009; Figlio, 1997).

For the ease of exposition, we have presented the two strands of literature as if they were relatively independent, while in practice, they are not. As Dohmen and Falk (2010) have clearly shown that the presence or absence of monetary incentives in the teaching profession induces the self-selection of different individuals.[3] Whether these two policy instruments are substitutive or complementary in nature is hard to judge because self-selection occurs based on unobservable characteristics, which in turn can be correlated to (unobservable) teacher quality. Merit pay wage policies should attract people who are expecting to benefit the most from such a scheme, but whether they are better able and/or greedier than average is difficult to gauge: as a consequence, it is almost impossible to predict what the overall effect on student achievement will be because the 'selection' and 'incentive' effects may work in opposite directions. If, therefore, it is impossible to derive uncontroversial predictions about what the most effective teacher policies are to improve school effectiveness, we do not have other alternatives than taking these questions to the data. In the next sections, we exploit cross-country and temporal variations in teacher recruiting and rewarding policies to identify which of them are the most effective in raising student achievement.

[3] '*…teachers are more risk averse than employees in other professions, indicating that relatively risk adverse individuals sort into teaching occupations under the current system. Using survey measures on trust and reciprocity, we find that teachers trust more and are less negatively reciprocal than other employees'* (Dohmen and Falk, 2010, p. F256).

3. Data and Descriptive Statistics

The growing interest in educational outcomes in recent decades has increased the demand for high-quality data from both academics and policy makers. International comparable surveys monitoring students' achievements at different grades have become more widespread in developed and developing countries. Namely, reading and literacy proficiency are tested in the fourth grade by PIRLS, numeracy proficiency is tested in two subsequent stages of the learning path – in the fourth and seventh grades – by the Trends International Mathematics and Science Study (TIMMS), while all three competencies (including scientific knowledge) are assessed at the tenth grade through the Programme for International Student Assessment (PISA). Although these surveys are informative from a cross-country perspective (Volante, 2017), they do not allow for a complete description of the learning environment for students. Information on teaching practices is self-reported by either the students or teachers/school principals, while only essential teacher characteristics are recorded, making it difficult to study the direct contribution of teachers to learning. Therefore, for our analysis, we combine survey data on students, teachers and schools belonging to the first cycle of the educational system by using a newly created data set of country-level time-varying institutional features relevant for primary teachers over the last century.

Microdata on students' achievements are drawn from the three waves of the PIRLS assessment, which was run every five years from 2001–2011 (the 2016 survey was not yet available in early 2017), covering about seventy country/state/region entities with legal autonomy in educational policy making. Information about the country-level institutional setting for teachers was collected from secondary data sources and assembled in a final data set that covers fifty-six countries/regions surveyed by PIRLS every year over the period from 1947–2011. Let us briefly describe the content of these data sets and present the basic descriptive statistics.

3.1 Microdata on pupil, teacher and school characteristics

The PIRLS assessment was conducted by the International Association for the Evaluation of Educational Achievement (IEA) and tested the reading literacy of fourth graders. The current study defines reading literacy as the ability to understand and use the written language forms required by society and/or valued by the individual. Three dimensions are assessed: the processes of comprehension, the purposes of reading and reading behaviours and attitudes. Student performance is measured by test scores in reading literacy, which are standardised to an international mean of 500 and a standard deviation of 100. The sample of countries is not balanced across waves, and some countries are missing in some waves. Internationally comparable test scores are provided for thirty-four countries/state/regions in 2001, forty-three in 2006 and fifty-eight in 2011: however, in Table 1, we report the list of countries included in our analysis, which here is restricted to countries/waves with

nonmissing values for any of the variables used in the empirical models (eighteen in 2001, forty in 2006 and forty-five in 2011).[4]

The test scores are nationally representative. National samples were drawn through a two-stage stratified sampling design. First, the participating schools were randomly selected. Then, within each school, a random sample of classes from the targeted grade was drawn and, within each class, all the students participated in the assessment. Together with students' reading achievement scores, the survey collected background information on students, parents, teachers, schools and curricular activities. The questionnaires were administered to the tested students, to their parents, to their reading teachers and to their school principals. The teachers' information refers to the main or unique reading teacher of the class; however, we ignore whether the same teacher taught the same students during the previous grades.[5]

Table 2 provides a summary of the statistics of the core variables used to perform the empirical analysis. Besides the reading test scores, we group the individual socio-demographic features and the school and teacher characteristics. We restricted the analysis to those students with a complete record of data for their parental background and for their schools and teachers. Among the individual observable features potentially responsible for the differences in performance, we consider gender, age in months at the date of the survey and language spoken at home. Socio-economic background is proxied by parental education and by a synthetic index for the available educational resources at home. The school features refer to the size, the geographical location or the presence of IT resources, as well as having a library for students. Among teacher characteristics, we focus on gender, age, educational level and tenure.

3.2 Institutional setting relevant for teachers

The PIRLS survey also collects some information about the institutional setting for primary school in each country and wave, which is provided by national country experts. We identify seven elements of a country educational system whose presence or absence could make the teaching profession more or less attractive, hence potentially affecting the average quality of aspiring teachers.

Namely, we selected the following dimensions as the most likely to affect teacher efficacy and student performance: (i) having a compulsory training period during (or immediately after) the teacher educational programmes required for teaching; (ii) passing a standardised test or an official examination as a basic requirement for teaching; (iii) having a compulsory probation period; (iv) the length of the possible probation period at the early stages of a

[4] We also replicate the analysis reported in the main text over the largest available sample without finding any significant differences. The results are available upon request.

[5] The structure of the data set is nested, with four levels of information aggregation: pupil – class and teacher – school – country. To have a perfectly nested sample, we dropped the very few (less than 1%) classes with more than one teacher of 'reading', while the inclusion of different classes with the same teacher is less harmful, unless one argues that teacher quality declines with the number of classes taught (this happens only in 758 classes over the 30,121 analysed).

teacher's career; (v) having a mentoring programme for teachers; (vi) having an official process to license or certify teachers by one of the following institutions: a teaching ministry, a national/state licensing board, a union or universities/colleges; and (vii) receiving a specific preparation on teaching techniques. For each dimension save for the fourth, we constructed a dichotomous indicator that takes a value of one if the requirement is present in a specific country at the date of the survey and zero otherwise; the length of the probation period instead is recorded in months. The process of granting credentials to teach can be the responsibility of different institutions, which is true as well of the evaluation process through certifications. Therefore, we also included a variable accounting for the total number of certifications required to teach at the fourth-grade level. Finally, we construct a synthetic *index of selectivity* as the mean of the previous dichotomous indicators for every country–year. A higher index means a more stringent selectivity for primary school teachers, which should translate into higher quality of the teaching staff and in higher student performance.

Because prevailing wages could affect the quality of the pool of teachers, we complemented the data on the reference institutional setting for primary schools with information on the corresponding average pay in each country. The average pay earned by primary school teachers is an indicator of the relative attractiveness of the profession compared with other professions that require similar qualifications in terms of education. Higher relative pay for teachers should attract better candidates (in terms of both observable and unobservable credentials) and/or enhance their quality in terms of their skills and motivation. Hence, from various issues of the OECD's *Education at a Glance*, we collect the ratio of primary teacher salaries to GDP per capita for each wave of PIRLS. Unfortunately, information on wages is available only for a subsample of thirty-one OECD countries. This reduced sample will be the smallest one on which we perform our empirical analysis.

Every country-level indicator is then associated with students by country/waves. A summary of the statistics of institutional variables are reported in Table 3 while in Table 4, we report the pairwise correlation matrix. Decomposing the standard deviation into its 'between' and 'within' components, a sufficient variation emerges within countries. Most indicators exhibit a positive correlation among each other, indicating that these dimensions complement and reinforce each other among educational systems. The few exceptions with a negative correlation are not statistically significant.

3.3 Reforming process of the teacher profession

We also collected detailed information about the reforming process that took place in each country over previous decades regarding the teacher profession, focusing on the selection and rewarding systems as the factors that affect the quality of teachers. Namely, we identified four reforming areas: (i) reforms affecting teacher recruitment processes; (ii) reforms on the pay scheme; (iii) reforms on teacher working conditions; and (iv) reforms affecting retirement possibilities. The first group of reforms refers to the *ex-ante* selection

process to become a primary school teacher, while the next three reform areas involve different dimensions of rewarding.

Among the reforms of the recruitment process, we included those that are changing the prerequisite criteria, through changes in the minimum marks to enter teacher colleges, in the level of educational attainment or in the prerequisites for teacher certification or licensing. Reforms of the working conditions refer to changes in the working hours, in the legal rights for special leaves or in continuous training. Reforms of pay include changes in the wage policy towards teachers, either as a part of a global civil servant reform or as a teacher-specific measure, which often comes from pressure from teacher unions. Finally, reforms of the retirement rules include legislative changes in the retirement entitlements (specific for teachers because we are interested in the incentive mechanisms of teachers' selection versus other careers) such as the prerequisites for early retirement and/or the level of pension benefits.

Information on the historical developments of the national education systems is drawn from different sources. The primary data source is the 'Database of National Labour, Social Security and Related Human Rights Legislation' (NATLEX) produced by the International Labor Office-ILO's International Labour Standards Department.[6] For the European countries, information has been cross-checked using *Eurybase*, the *Eurydice* database that provides detailed information on European education systems and policies since the end of World War II. Moreover, we also use country-specific descriptions of national education systems and thematic studies on specific institutional features. Finally, we double-checked our data by directly contacting national experts in the field.[7]

Exploiting these data sources, we could identify the exact year of implementation of each reform, as well as the direction of the change that took place (i.e., whether it was favourable or not for teachers). As a result, we assembled an original data set containing yearly observations for the period from 1947–2011 for all countries surveyed in PIRLS on the implementation of legislative changes in the previous four dimensions of the teaching profession. Whenever, in a given year, we recorded a change in a specific dimension, we assigned a value of one, but if no changes occur, we assigned a value of zero. When legislators have repeatedly reformed a specific dimension over the sample period, we created step dummies, which were then summed up over the years, with a final normalisation yielding a unitary range of variation. For all the dimensions, we constructed the indicators based only on the direction of the legislative change so that an increase or

[6] The database lists and classifies all the legislative actions in several fields that are broadly related to the labour market and working conditions, ranging, among others, from employment security to maternity protection or the elimination of forced labour. We focused on the categories 'education, vocational guidance and training', 'conditions of employment', 'conditions of work' and 'specific categories of workers => teachers'. For each action, several pieces of information are provided: name, country, type of legislation, adoption, entry into force, publication date, ISN code, an abstract, a short description and – when applicable – links to related texts (i.e., basic texts, repealing texts, repealed texts, amended texts and so on). Among all the legislative actions recorded, we selected those relevant to our scope and classified them according to the four broad categories described in the text.

[7] Although comprehensive and constantly updated, the NATLEX database could unintentionally misreport or omit some legislative act or regulation. Symmetrically, specific collective agreements regulating contracts in the private sector are not recorded. However, it collects legislation actions with *erga omnes* effects, which are more interesting for our purpose. Finally, NATLEX may not report wage adjustments not requiring an explicit normative act, such as price indexing. A measurement error could arise from these limitations, creating a downward bias in our estimates.

decrease in the variable refers to a legislative change that is favourable or unfavourable to teachers. We ended up with four indicators. The first index refers to the selectivity of teacher recruitment: an increase corresponds to more restrictive selection criteria. The second one is related to working conditions, and an increase refers to a reform allowing for more favourable working conditions (workload, holidays, standard requirements and the like). The third indicator is defined according to changes in the wage policy and salary conditions; also in this case, an increase means more generous wage allowances for primary school teachers. Finally, the fourth one captures the stringency of retirement conditions and includes the retirement allowance, severance pay and retirement age. The indicator increases whenever retirement conditions are more favourable. The time plot of these variables is reported in Figure 1, while the original timing of the reforms is reported in section A.2 of the Appendix.

We then match these indicators to the teachers surveyed by PIRLS according to the year when they entered into the labour market. Consider the case of 'teacher recruitment reform indicator' in Italy as an example. The sources for the data report more stringent reforms for this dimension in 1987, 1988 and 1990. Therefore, we constructed a variable that is zero before 1987, one-third in 1987, two-thirds in 1988 and 1989 and one afterward. Every teacher entering the profession before 1987 gets a zero value for this reform, those entering in 1987 get one-third, those in 1988–1989 receive two-thirds and a value of one for teachers hired more recently.

Table 5 reports the summary statistics of our original reform variables not yet matched with students' and teachers' PIRLS microdata, while in Table 6, we report the pairwise correlation matrix among them. Notice that the decomposition of the standard deviation into the 'between' and the 'within' components indicates that there is a sufficient variation within countries.

4. Empirical Strategy

The aim of the empirical strategy is to identify whether some policies intended to *attract, select* and/or *motivate* good teachers who can improve student performance in primary schools. To test whether teachers matter for student performance, one should correlate student achievement with measures of teacher quality. However, we have already reviewed the problem of measuring teacher quality: the observable characteristics of teachers are weakly correlated with student achievement, and the reverse strategy of inferring teacher quality from observed student achievements is only valid when either the students are randomly allocated to teachers (inapplicable for countries where there is explicit or implicit streaming) or one possesses longitudinal samples where repeated observations of different student cohorts are exposed to the same teacher (Rivkin et al., 2005). This strategy is even more complicated when we consider that students are often exposed to more than one teacher (a sort of *group production*) and that teacher mobility is often driven by perceived student teachability (thus inducing a *self-sorting* of teachers to schools/classes). Given the repeated cross-sectional structure of the data available in PIRLS, we exploit both cross-sectional and temporal variations to identify the policies that may be effective either because they attract or select better teachers or because they solicit a higher level of effort. In addition, we focus only on the effect of the main instructor in charge of teaching

reading to fourth-grade students. Unfortunately, the survey does not collect information about teachers to whom students have been exposed to in previous grades, if different from the current one, making it impossible to distinguish between the two different effects.

The empirical analysis proceeded in three subsequent steps. We first reviewed the standard correlations of student achievement with individual, class and school characteristics, as well as the characteristics of the teachers responsible for reading. For each pupil i associated with class/teacher j in school s of country c surveyed in year t, we estimated a standard educational production function for the student's reading achievement a_i through the following equation:

$$a_{ijsct} = \alpha + \beta X_{it} + \gamma Y_{jt} + \eta Z_{st} + \delta_t + \delta_c + \varepsilon_{ijsct} \quad (1)$$

where the vector X_{it} is associated with the students and contains information about gender, age in months, language spoken at home, immigrant status, parental education and available educational resources. The vector $Y_{jt} = [X_{-ijt}, T_{j\tau t}]$ associated with the class can be decomposed into two subvectors: the first one, X_{-ijt} includes the contextual class effects computed when excluding the considered pupil (like the share of females in the class, average age in months, share of immigrants and/or of students speaking a different language at home, an index for household educational resources, average educational attainments among the parents in the class, etc,); the second subvector $T_{j\tau t}$ contains information regarding the main or unique reading teacher of class j, who entered the labour market in year τ and was surveyed in year t: gender, age (in ten-year intervals), tenure (years) and educational attainment (being a graduate and having an official certification for teaching). The third vector Z_{st} includes the school characteristics such as location (urban/rural), average teacher tenure in the school, availability of a library and IT technologies, share of disadvantaged students in the school and school size. δ_t and δ_c are wave and country-fixed effects, while the idiosyncratic error component, clustered at the class level, is ε_{ijsct}. School-fixed effects can also be considered instead of including school characteristics, and in this case, the country-fixed effects are removed.

The second step consists of the analysis of the contextual effects affecting primary school teachers at the country level and that can possibly change from one survey to the other. Hence, we introduced institutional features and policies targeted at teachers and teaching activities. We estimate the following extended model:

$$a_{ijsct} = \alpha + \beta X_{it} + \gamma Y_{jt} + \eta Z_{st} + \sigma W_{ct} + \delta_t + \varepsilon_{ijsct} \quad (2)$$

where we augment the previous specification with the vector W_{ct}, which contains information about the institutional design of the teacher recruitment process at the country–year level, as well as their relative wage. Obviously, this specification has two main limitations. First, because institutional features are observed only over one decade, they show limited (if any) variation within a country over time. Hence, the estimated effects might be confounded with other institutional elements at the country level, which are not explicitly accounted for and prevent the inclusion of country-fixed effects in the model. Second, because we do not have retrospective information on these characteristics, they cannot be used to analyse the teachers' job market features at the time of entry into the profession. We can only estimate the effect of a contemporaneous correlation between the

relevant settings for teacher quality and student performance. As such, they are likely to directly affect only the quality of the pool of aspiring teachers and – through positive externalities – the overall quality of the teaching body.

Finally through the following equation:

$$a_{ijsct} = \alpha + \beta X_{it} + \gamma Y_{jt} + \eta Z_{st} + \sigma R_{c\tau t} + \delta_t + \delta_c + \varepsilon_{ijsct} \quad (3)$$

we exploit exogenous variations in the labour market setting for teachers prevailing in the year of their entry in the labour, as measured by the vector $R_{c\tau t}$ of the implemented reforms in country c at time t. The match of the reforms to teachers according to their year of entry into the labour market allows for identifying the effect of policies by comparing students' achievement in classes taught by 'treated' teachers against classes taught by 'nontreated' teachers acting as control cases. In fact, by matching the reforms to teachers based on their age and experience, we can distinguish those who were affected by the reforms those who were not. For example, suppose a reform introducing the requirement of a university degree (BA level) to become teacher was approved in a country in 1990. As a consequence, candidates leaving teaching schools in the same year were forced to undertake three additional years of college to obtain the degree. Thus, all other things constant, we can test whether the students in classes with teachers hired before 1990 exhibited worse performance compared with those taught by teachers hired after 1990 (presumably with a BA degree – because of a lack of information, we are forced to assume perfect compliance). In addition, in the present case, the effect of the reforms is more precisely identified because the age and tenure effects are separated by observing individuals in the same labour market with the same age and tenure but matched with different sets of teacher policies because they are observed in different time periods (thanks to the availability of the three surveys that span a decade). Finally, the repeated cross-section nature of our data allows us to distinguish the age effect and the cohort effect.

Our identifying assumption is that reforms in teacher policies (especially recruitment ones) affect beginner teachers only, leaving already tenured teachers unaffected (i.e., any imitative behaviour can be considered negligible). Similarly, we consider that reforms regarding pay and retirement rules affect all teachers but at a different degree of intensity, here being stronger the younger that the teacher is (i.e., the smaller the time period between entry into the profession and the reform is, via an effect on the attractiveness of the profession).

Furthermore, the lack of detailed information on each reform makes it impossible to construct a quantitative measure of the effect, allowing for a comparison of the magnitude of their impact across countries and over time. As such, our variables capture the frequency and intensity of the reforming activity of subsequent governments *vis-à-vis* the teachers within each country.[8]

In the following section, we estimate the models corresponding to equations (1)–(3), including some variants to check their robustness. The list of countries and waves included in the three data sources does not perfectly overlap. To have a consistent sample across all specifications, we rely on fifty-six countries for which we have complete information for all

[8] A similar strategy has been pursued by Braga et al. (2013) while studying the impact of educational policies on educational inequalities.

the variables included in the main model, except wage, which is available only for a subset of OECD countries. We run our analysis fully and with a restricted sample to test whether the results are sensitive to a restriction of the rich/OECD countries sample.

5. Results

In this section, we present the main results in gradual steps. Estimates for the educational production function described in equation (1) are presented in Table 7, where we correlate student performance with the teachers' characteristics, controlling for the students' demographics, socio-economic backgrounds and class and school characteristics. We use alternative specifications, including school-fixed effects (column 1), school characteristics (column 2) or country-fixed effects (column 3). Furthermore, for the sake of comparability with the following models, columns (4)–(6) replicate the same specifications on the subsample of countries for which the data on teachers' salary are available, that is, OECD countries. In all the specifications, the standard errors are clustered at the school level.

Regardless of the specification, in line with previous evidence, female students outperform their male counterparts by an average of twelve points. The point estimates show a very small negative effect of age that probably captures the lower skills of students repeating the year. The socio-economic background of students is positively correlated with their attainment: students with more educated parents and/or better educational resources available at home obtain higher scores in reading. Moreover, students speaking a different language at home are at a disadvantage. It is important to note that the PIRLS survey directly tests linguistic competencies that are extremely correlated with the language usually spoken in everyday life. Interestingly, the same characteristics averaged by class play the same role in determining pupil performance and strengthening the effect of the corresponding individual feature, indicating a significant peer effect.

Moving to the core of our research on teachers, we find that some observable characteristics of the prevalent teacher in the class are statistically correlated to student performance. The signs of gender, age and tenure are in line with previous studies (Clotfelter et al., 2007; Croninger et al., 2007). On average, female and younger teachers exert a positive effect on literacy. The gender effect is robust, irrespective of the sample and specification. However, the age effect is sensitive to the chosen specification, and it dissipates in the restricted sample only when school- or country-fixed effects are accounted for. After considering the teacher's age, tenure (in terms of years of continuous activity in primary school) does not always influence student performance.[9] Instead, different from other studies, we find a statistically significant effect of teacher education (Chingos and Peterson, 2011). In particular, having a teacher with at least a tertiary level of education increases, on average, student performance on the standardised tests by four to seven points, depending on the specification. Increasing teachers' educational attainment seems to be a valuable driver to enhance primary student performance. The same effect has not been found in previous studies, which have mainly focused on higher levels of the educational system. However, this result is not surprising because in the past, teachers

[9] This is important from our perspective because we match teachers and reforms based on this variable. Its limited significance in this regression reduces the risk of a spurious correlation with the reform variables.

entered primary schools after completing teaching schools, corresponding only to a secondary or vocational level of education. The completion of a tertiary level of education has typically been required for higher school levels. Over time, the requirements to enter the teaching profession have increased, and hence, on average, a differential effect emerges between teachers with a stronger educational background and those with a weaker one. However, when including school characteristics instead of school-fixed effects, the coefficient drops and becomes negative and significant in the sample of OECD countries, where the teacher profession is often less attractive for more educated and qualified new entrants into the labour market.[10]

Having identified the basic determinants of student performance, we move ahead in our analysis by exploring the role of country-specific institutional features in shaping the average test scores, as described by equation (2). The results are presented in Table 8. In all the specifications, we control for the same variables – at the individual, teacher, class and school level – included in previous estimates. School-fixed effects cannot be considered in the model because institutional features are country invariant in any survey year. In column (1), we report the results of distinct regressions where each institutional feature is included alone, while in column (2), we report the coefficient of an overall index of selectivity, which is obtained as a normalised mean of the previous eight indicators. Column (3) replicates the specification of column (1) but here with the inclusion of a pay index. Columns (4) and (5) include the same models as in columns (1) and (2) but are restricted to the subsample of OECD countries, while the remaining columns show the effects of pay alone (6) and the interaction between pay and selectivity (7).

All the institutional characteristics included in the model are positively correlated to student performance. Countries characterised by more selective recruitment or better-structured training periods enhance student competencies. The dimensions with a stronger effect are those related to the training process, especially when constituted by an official training period before (or immediately after) entering the teaching profession. A similar role is exerted by the completion of a probationary period. More specifically, in countries where teachers receive a specific preparation on how to teach reading, in the full sample, students score twenty-nine additional points more than the average (eighteen in the restricted one). Similarly, in countries envisaging a probation period for primary school teachers, the average reading test scores are from ten to seventeen points higher than elsewhere. Finally, having to pass an examination before starting one's teaching career is also positively related to student achievement, with an average effect ranging between five and seven points. The certification process is the only institutional feature that is sensitive to the estimation sample: having an official certification process to license primary school teachers and increasing the number of required certifications is positively associated with student performance in the full sample (including middle-to-low income countries), while it reverses its sign when the sample is restricted to high-income countries. This could be taken as indicative that simple accreditation, without adequate training, does not produce an increase in the quality of teaching, especially where the supply of potential teachers is relatively abundant.

[10] To see whether country-specific heterogeneity drives these results, we estimate equation (1) for each country. The Netherlands and Quebec are the two countries/regions that account for the negative effect of a tertiary level of education on student performance. Instead, the age effect is heterogeneous across countries.

As a further investigation, we run a regression where all the considered institutional features are accounted for in a synthetic way through the index of selectivity to measure the degree of selectivity and specialisation for primary teachers in a country. This specification allows for testing whether our previous estimation captures some other country-specific effect, not only the institutional setting of reference for primary teachers. The results presented in columns (2) and (5) corroborate, in both samples, with those obtained when each single dimension is taken into account. Overall, on average, students perform better in countries where the entry in the teacher profession is more selective and training specialisation is more intensive than elsewhere.

In column (6), we analyse the effects exerted by remuneration, including as a regressor the ratio between the wages earned by teachers and the country GDP per capita, as an indicator of the relative attractiveness of the profession compared with others. A higher ratio should enhance teacher quality both in terms of skills and motivation. Indeed, the coefficient that is associated with the pay index is positive and significant, showing a strong positive correlation between the reward of teachers and the performance of their students (which is in line with the findings of Dolton and Marcenaro-Gutierrez, 2011). Relative remuneration appears as one of the institutional features that enhances student performance by attracting better candidates among possible entrants into the teaching profession, but also as a tool to retain and motivate them throughout their careers.

Finally, in column (7), we estimate the same model with the inclusion of an interaction term between the index of selectivity and the pay ratio. Teacher selection and reward appear as substitutes; that is, the effect of wage on student performance is attenuated by an increase in selectivity, and the effect of selectivity on performance declines as wage increases. From a policy perspective, because both instruments are positively correlated with primary students' attainment, the choice between the two alternatives should be based on a cost-effectiveness criterion.

However, any conclusion based on time-invariant institutional features is exposed to the risk of a spurious correlation, even though some of these dimensions change over the different survey years. Mainly for this reason, we resort to our preferred strategy, which is represented by the model described by equation (3). In Tables 9 and 10, we test this hypothesis, considering the effect on student performance from a change in the institutional setting relevant for primary school teachers just before starting their careers. All the models are estimated, including student, class, teacher and school characteristics together with wave and country-fixed effects (Table 9) or without country-fixed effects (Table 10). In the first three specifications, we consider all the reforming dimensions together, while in the last three specifications, we run different regressions for each single reform. As before, we conduct the analysis considering the whole sample of countries and the subsample of countries where information on remuneration was present to control for the effect exerted by the pay level. The former are presented in columns (1) and (4), and the latter are presented in the remaining columns.

The two reforming areas potentially enhancing student performance through teacher quality are the introduction of a more selective recruitment process and more favourable reward policies, a sort of 'warm glove' attitude towards teachers, which includes more

generous pay conditions, more favourable working conditions and/or more advantageous retirement policies. A more selective or targeted recruitment process is implemented to select well-qualified candidates who have specific skills. Symmetrically, changes in the reward scheme should improve the quality of the teaching workforce by attracting more qualified and motivated candidates, reducing turnover, increasing retention and, hence, enhancing students' achievements. Although the nature of our reform variable does not allow for distinguishing between a simple revision in the wage structure or the introduction of a more sophisticated pay-to-performance scheme, our results indicate that the level of stipend paid to teachers is a possible lever for policy makers.

The selectivity of the recruitment policy is robust to the sample, as well as to the inclusion of country-fixed effects, when accounting for unobservable heterogeneity within countries over time. Also, when considered together with other reforms, our proxy for selectivity still affects student performance (column 1), even when controlling for the wage level (column 2). Instead, the effect of the remuneration reforms is sensitive to the chosen specification; they exert a positive and significant effect only when country-fixed effects are not included in the specification (Table 10). On the contrary, having a teacher who enters the labour market just after an improvement in working conditions decreases, on average, the performance of their students, signalling that such policies tend to attract less-motivated individuals or those with conciliation problems because of housewiving duties (e.g., female teachers who find this profession easier to combine with caregiving in their households). A negative effect is found when considering changes in the retirement schemes (for the OECD restricted sample, when including country-fixed effects), indicating again that the possibility of early retirement may lower the level of individual motivation, possibly because of short-sightedness. The results hold irrespectively when controlling or not controlling for the pay level of primary school teachers. Overall, the results confirm that selection at entry is as good as improving pay conditions when it comes to raising student performance. Unfortunately, our reform variables are scale-free, making it impossible to assess the size of the existing trade-off between the two alternatives.

5.1 Further analysis

As a further step, we also investigate whether the institutional setting has a differential effect depending on teacher characteristics. We then interact each of our institutional features with the dichotomous indicator for having a teacher with a tertiary level or post-tertiary level of education. The estimates are presented in Table A.1.1 in the Appendix, where in each line, we report the point estimates for the distinct regressions where each institution is included alone, controlling for all the available observables. In this specification, the estimated coefficient of the institutional feature is the average effect on nongraduate teachers, while the effect for graduate teachers will be the sum of the estimated coefficient for the institutional feature and the interaction with the dummy for holding a graduate degree. In particular, the marginal effects of each institutional feature are reported for graduate teachers in column (1), for nongraduate teachers in column (2) and their difference in column (3). In columns (4)–(6), we replicate the same structure on the restricted sample for which information on pay is available. Though the effects are not the

same along all the dimensions in terms of magnitude, the overall picture is that these features are more effective in enhancing the teaching quality of more educated instructors. Better-educated teachers are associated with better student performance in countries where the institutional setting is more selective. In the subsample of OECD countries when controlling for the pay level, a slightly different situation manifests. In rich countries, the overall index for the institutional framework seems to be more effective for less-qualified teachers, indicating that part of the selection and/or the attractiveness is exerted by attending university courses. In addition, in this case, the results differ along the considered dimensions. The results are robust to the sample, as well as to the inclusion of the pay index.

We also explore the heterogeneity by interacting each reform with the level of education of the teacher. The estimates are presented in Table A.1.2 and show a clear trend: the effects of all reforms are stronger for nongraduate teachers in OECD countries. In more detail, stricter recruitment processes have a positive effect that is always significantly larger for nongraduate teachers, while retirement reforms have a positive effect for nongraduate and a negative effect for graduate teachers. Pay reforms are effective in improving the performances of (children taught by) both graduate and nongraduate teachers, with the latter overperforming compared with the former in OECD countries and the opposite in the whole sample. Reforms for working conditions affect nongraduate teachers less negatively than graduate ones.

In all previous specifications, we have focused on the intensity of the reform processes that occurred before each teacher entered the labour market; these reforms should affect the quality of the applicants and, hence, the subsequent performance of their students. However, having been exposed to reforms throughout their careers could influence the incentives to be effective in teaching. Therefore, in Table A.1.3, we study whether the intensity of the reform process throughout one's career has an effect on teacher quality and translates into different levels of student achievement. In particular, for each of the four reforming areas, we identify the number of legislative changes affecting a given teacher after his or her entry into the school system up to the date of the survey when the students' competencies are tested. All the models are estimated, including student, teacher, class and school characteristics, together with wave and country-fixed effects. Teachers exposed to more changes in the recruiting process or in pay conditions are more effective in improving student performance. The pay structure and working conditions appear as two effective tools to influence the incentives for primary school teachers (though in an opposite direction). On the contrary, a higher number of reforms in the retirement process have no clear effects on student performance, indicating that these changes do possibly change incentives throughout a teacher's career.

An implicit and untestable assumption of our identification is that selecting the teaching profession is more affected by the most recent reforms rather than by reforms targeting older cohorts (i.e., we assume a decreasing marginal salience of reforms over time). This assumption is necessary to identify the effect of reforms when using individual-level information. Hence, we test whether the effects of the institutional setting dissipate over time by running a separate analysis according to teacher tenure. As shown in Table A.1.4, based on the synthetic index for the selection process, enhancing selectivity is more effective for young teachers. Similarly, according to the results in Table A.1.5, the effect of

relative pay (proxied by the salary/GDP per capita ratio) tends to dissipate with tenure, being more intense for shorter-tenured teachers. Finally, Table A.1.6 shows that the reform activity has a heterogeneous effect according to tenure. In particular, although reforming working conditions and retirement rules have a negative effect that is mostly homogenous over tenure length, the effects of recruitment reforms and salary reforms raise their intensity for longer-tenured teachers than for others.

In addition, to better characterise the role played by pay level in attracting better workers as primary school teachers, we found information about the (1) ratio of salary at the top of the scale compared with the starting salary, (2) years elapsing from rising from a starting salary to a top salary and (3) salary per hour of net contact (teaching) time after fifteen years of experience. Although not fully comparable because the data are available only for a subset of countries, the estimates are in line with those discussed in the main text. In detail, the sign and significance of the top/least salary ratio is the same as our main indicator, which is positive and always significant. The steepness of the salary curve is positively correlated to the performance of students, but the significance disappears once we control for the top/bottom wage ratio, indicating that what matters is the overall salary profile more so than how fast it takes to reach the top. Finally, the hourly wage is negatively correlated to student outcome, but in this case, the result is not significant once we control for the total salary, indicating that there is a negative correlation between hourly wage and hours worked and that student outcomes increase with total teacher income rather than with hourly wage. Stated differently, keeping the hourly wage fixed, the effectiveness of teachers increasing is positively correlated to their hours worked.

Finally, though not reported in the text, we also collected data on the statutory teacher wage at the start of their career, the average wage after ten and fifteen years of experience and the wage at top of the pay scale. To match the wages and teachers at different seniority levels, we linearly interpolate the four points on the wage scale, assuming then that wages smoothly increase over time. The likely introduction of a measurement error regarding the true pay received by each teacher does not bias the other coefficients, and at worse, it implies an attenuation bias. Nevertheless, it allows a rough comparison of pay scale steepness across countries and over time. We run our model (2) excluding age from the controls, and the results indicate that a 1% increase in wage translates into a three points higher test score, which is 0.6% of the standardised average score.

6. Robustness Checks

To corroborate the core results of our analysis, in this section, we perform a series of robustness checks.

One possible concern for our identification is the nonrandom allocation of institutions and reforms over time and across countries. We therefore perform a falsification test in which we estimate our basic equations (equations 2 and 3) using as the regressors a vector of randomly generated institutional features prevailing in the three waves or when teachers enter the labour market. The estimated coefficients of the falsification tests for the institutional setting are presented in Table A.1.7, where we randomly assign institutions across countries and survey years. The point estimates for both the single institutional

features and for the synthetic indices are no longer statistically significant, and when they are significant, they obtain the opposite sign (with the only exception being the examination dummy). Overall, these falsification tests indicate that our original regressions actually capture an effective association of changes in the institutional setting and do not contradict a causal interpretation of the results.

Symmetrically, we randomly allocate the reforms to teachers in two different ways. First, we randomly assign the reforms across countries within tenure cohorts; second, we assign the reforms to teachers randomly both across countries and tenure cohorts. As shown in Table A.1.8, also in this case, the estimated coefficients are no longer significant.

As a second sensitivity test, to check whether the results are driven by the behaviour of a single country, we re-estimate our baseline regressions, here excluding from the sample one country at a time or for each country separately. Although not reported in the current paper because of space limitations, both for institutional features and reforming activity, the results continue to hold and are not driven by the behaviour of a single country.[11] Similarly, we perform separate regressions for different geographical areas to test whether specific groups of countries are driving our results. In particular, we perform separate regressions for Europe, formerly planned economies, North America, Latin America, East Asia, the Middle East and North Africa, Oceania and sub-Saharan Africa. No specific patterns emerge when using regional disaggregation.

7. Discussion and Conclusions

The current paper provides new evidence on the effect of teacher quality on student performance in primary school. Based on international standardised tests for literacy conducted with fourth-grade students and using variations in the institutional setting, our analysis shows that teacher quality matters. The two main channels for enhancing teacher quality are selectivity when the potential teachers are entering the profession and the rewards they are given when teaching. Here, reward can be disaggregated into compensation, working conditions and early retirement eligibility, but they emerge as statistically significant although with an opposite sign: offering higher pay or a lighter working arrangement (possibly accompanied by advantageous retirement conditions) seem to work in opposite directions when it comes to attracting high-quality teachers. However, it is important to recall that teachers' salaries represent the largest cost in providing school education. Therefore, from a policy perspective, setting an adequate level of pay is essential to ensure both quality teaching and balanced educational budgets.

Our results indicate that policies could effectively enhance school quality. Indeed, some of the dimensions we consider are actually introduced by policy makers to better train teachers and provide them with adequate skills. Better formal professional training shows up as a key factor in being successful and effective in class teaching practices, improving learning and ensuring high-quality standards. Specifically, on the one hand, the presence of examinations or licences is a screening device to assess the competencies and capabilities of aspiring teachers. On the other hand, the purpose of a probationary period is to ascertain

[11] The results are available upon request.

whether the conduct and work performance of the teachers meet the standards expected by their employers before the full rights and responsibilities of ongoing employment are confirmed. The probationary period also provides support and feedback to teachers at the beginning of their careers. During this period of supervision, it is possible to closely evaluate the advancements of newly hired workers. Finally, the initial training ensures that all teachers possess adequate knowledge, attitudes and resources to perform well. Effective preservice or in-service induction systems should translate into a more homogenous pool of highly qualified teachers delivering high-quality education.

We have not limited our investigation to the prevailing institutional framework in each country (thus, mainly exploiting cross-country variability), but we have also analysed the reforming activities of government in the four main areas of human resources management: recruitment, working conditions, pay and retirement conditions. Our identifying assumption consists of reforms mostly affecting teachers at the time of their entry into the profession and gradually dissipating afterwards. As such, these reforms are salient in attracting better-qualified and more motivated candidates. We find that selection and – to a lesser extent – generous pay are both effective tools to attract good aspiring teachers, but when jointly considered, they appear as substitutes because each of them reduces the effectiveness of the other. Thus, reforming the selection process or the reward scheme must be adequately balanced if policy makers want to improve student performance. In addition, when targeted at primary school teachers, these reforms could also be effective in enhancing the overall quality of the educational system thanks to their cumulative effect on subsequent school grades.

It is important to recall that our previous estimates do not capture all aspects of a country setting that might crucially affect teachers' incentives. In some countries, for example, stipends are only a fraction of total remuneration, which will sometimes include health insurance, pensions or fringe benefits. In these cases, wage differentials over time and across countries capture only one dimension of the rewarding scheme that could influence teacher productivity. Our analysis also neglects other dimensions of the educational process, ranging from school infrastructure to extra-curricular activities or school timetables. These dimensions, whether proxied or not by our contextual controls, could be responsible for the heterogeneity of the reforming activities, which we have documented in our extended analysis.

References

Andersson, C., Johansson, P. and Waldenström, N. (2011). Do you want your child to have a certified teacher? *Economics of Education Review*, 30(1): 65-78.

Arias, J.J. and Scafidi, B. (2009). When Does Teacher Licensure Make Sense? *The B.E. Journal of Economic Analysis & Policy*, 9(1), 4.

Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H. and Wilson, D. (2009) Evaluating the impact of performance-related pay for teachers in England. *Labour Economics*, 16 (3): 251-261.

Bacolod, M.P. (2007). Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality. *The Review of Economics and Statistics*, 89(4): 737-751.

Ballou, D. (2001). Pay for performance in public and private schools. *Economics of Education Review,* 20: 51–61.

Boarini, R. and Lüdemann, E. (2009). The role of teacher compensation and selected accountability policies for learning outcomes: An empirical analysis for OECD countries. *OECD Journal: Economic Studies*, 2009(1):1-20.

Braga, M., Checchi D and Meschi E. 2013. Institutional Reforms and Educational Attainment in Europe: A long run perspective, *Economic Policy* 73/2013: 45-100.

Buddin, R. and Zamarro, G. (2009). Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics*, 66(2): 103–115.

Chetty Raj, John N. Friedman and Jonah E. Rockoff, (2014a). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates, *American Economic Review*, vol. 104(9): 2593-2632.

Chetty Raj, John N. Friedman and Jonah E. Rockoff, (2014b). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, vol. 104(9): 2633-2679

Chingos, M.M. and Peterson, P.E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, 30: 449–465.

Clotfelter, C.T., Ladd, H.F. and Vigdor, J.L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, 26: 673–682.

Croninger, R.G., King Rice, J., Rathbun, A. and Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, 26(3): 312–324.

Dohmen, Thomas and Armin Falk. 2010. You get what you pay for: incentives and selection in the education system. *Economic Journal* 120: F256-271

Dolton, P. and Marcenaro-Gutierrez, O.D. (2011). If You Pay Peanuts do You Get Monkeys? A Cross Country Comparison of Teacher Pay and Pupil Performance. *Economic Policy*, 1: 5–55.

Dolton, P., Tremayne, A. and Chung, T.-P. (2003). The Economic Cycle and Teacher Supply. A paper commissioned by the Education and Training Policy Division, OECD, for the Activity *Attracting, Developing and Retaining Effective Teachers* (March 2003).

Falch, T., Johansen, K. and Strøm, B. (2009). Teacher shortages and the business cycle. *Labour Economics*, 16(6): 648-658.

Figlio, D.N. (1997). Teacher salaries and teacher quality. *Economics Letters*, 55: 267–271.

Garrouste, C. (2010): *100 years of educational reforms in Europe: a contextual database.* EUR – Scientific and Technical Research series, Vol. 24487, (2010): pp. 1-349.

Goldhaber, D., Brewer, D., and Anderson, D. (1999). A Three-Way Error Components Analysis of Educational Productivity. *Education Economics*, 7(3), 199-208.

Hanushek, E.A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30: 466–479.

Hanushek, E.A. and Rivkin, S.G. (2006). Teacher Quality, in E.A. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education Volume 2*. Amsterdam: Elsevier, Chapter 18, pp. 1051–1078.

Harris, D.N. and Sass, T.R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics,* 95: 798–812.

Holanda, M.C., Nogueira, C.A.G. and Petterini, F.C. (2008). Education regulation as an instrument of income distribution: The case of Ceara. *The Quarterly Review of Economics and Finance*, 48(2): 377-388.

Kane, T.J., Rockoff, J.E. and Staiger, D.O. (2008). What Does Certification Tell us About Teacher Effectiveness? Evidence from New York City. *Economics of Education Review,* 27(6): 615-631.

Lavy, V. 2015. Teachers' pay for performance in the long-run: effects on students' educational and labor market outcomes in adulthood. NBER wp n.20983

Lazear, E. and M. Gibbs. 2007. *Personnel Economics for Managers* (2nd edition). Wiley

Moor Johnson, S. (2006). The Workplace Matters: Teacher Quality, Retention, and Effectiveness. Best Practices-NEA Research Working Paper (July 2006).

Nagler, Markus, Marc Piopiunik and Martin R. West. 2015. Weak markets, strong teachers: recession at career start and teacher effectiveness. NBER wp n.21393

OECD (2018), Education at a Glance 2018: OECD Indicators, OECD Publishing, Paris

Ortega, D.E. (2010). The effect of wage compression and alternative labor market opportunities on teacher quality in Venezuela. *Economics of Education Review*, 29(5): 760–771.

Pelayo, I. and Brewer, D.J. (2010). Teacher Quality in Education Production, in D.J. Brewer and P.J. McEwan (eds.), *Economics of Education*. Oxford: Elsevier: 178-182.

Rivkin, S., E.Hanushek and J.Kain. (2005). Teachers, Schools, and Academic Achievement, *Econometrica*, vol. 73(2): 417-458

Santibañez, L. (2006). Why We Should Care if Teachers Get A's: Teacher Test Scores and Student Achievement in Mexico. *Economics of Education Review* 25 (15): 510-520.

Schächter, J. and Thum, Y.M. (2004). Paying for high- and low-quality teaching. *Economics of Education Review*, 23: 411–430.

Volante, L. (ed) (2017). *The PISA Effect on Global Educational Governance*. Routledge Press.

Watlington, E., Shockley, R., Guglielmino and P. Felsher, R. (2010), The High Cost of Leaving: An Analysis of the Cost of Teacher Turnover, *Journal of Education Finance*, vol. 36 (1): 22-37.

TABLES

Table 1: List of data availability by countries/regions.

	Main sample	PIRLS institutions and reforms				OECD
		2001	2006	2011	Total	salary
1	Australia	n/a	n/a	1696	1696	yes
2	Austria	n/a	3316	3655	6971	yes
3	Azerbaijan	n/a	n/a	3065	3065	
4	Belgium (Flemish)	n/a	3261	n/a	3261	yes
5	Belgium (French)	n/a	2422	2114	4536	yes
6	Bulgaria	n/a	2766	4581	7347	
7	Canada, Alberta	n/a	2557	1398	3954	yes
8	Canada, British Columbia	n/a	1950	n/a	1950	
9	Canada, Nova Scotia	n/a	2852	n/a	2852	
10	Canada, Ontario	2620	2525	2774	7919	yes
11	Canada, Quebec	2029	2072	2721	6822	yes
12	Chinese Taipei	n/a	n/a	3543	3543	
13	Croatia	n/a	n/a	3833	3833	
14	Cyprus	1018	n/a	n/a	1018	
15	Czech Republic	n/a	n/a	436	4117	yes
16	Denmark	n/a	2600	3106	5706	yes
17	England	1381	1234	n/a	2615	yes
18	Finland	n/a	n/a	3408	3408	yes
19	France	n/a	2662	3065	5727	yes
20	Georgia	n/a	2338	3353	5691	
21	Germany	n/a	3710	1942	5652	yes
22	Honduras	n/a	n/a	988	988	
23	Hong Kong SAR	3708	3623	2666	9997	
24	Hungary	3300	2809	4081	10190	yes
25	Iceland	n/a	1414	n/a	1414	yes
26	Indonesia	n/a	2593	1555	4148	yes
27	Iran	4987	3000	4454	12441	
28	Ireland	n/a	n/a	3256	3256	yes
29	Israel	n/a	1337	2212	3549	yes
30	Italy	3108	2447	2723	8278	yes
31	Kuwait	n/a	866	629	1495	
32	Latvia	1956	2962	n/a	4918	
33	Lithuania	1821	3605	3925	9351	
34	Macedonia	n/a	1561	n/a	1561	
35	Malta	n/a	n/a	336	336	
36	Moldova	1450	2975	n/a	4425	
37	Morocco	n/a	1306	4684	5989	
38	Netherlands	1587	1775	1539	4901	yes
39	New Zealand	1554	2818	2557	6929	yes
40	Northern Ireland	n/a	n/a	1253	1253	
41	Norway	2193	2064	1456	5713	yes
42	Oman	n/a	n/a	4773	4773	
43	Poland	n/a	3497	3925	7422	yes
44	Portugal	n/a	n/a	3158	3158	yes
45	Qatar	n/a	1597	1597	3194	
46	Romania	n/a	3042	3636	6678	
47	Russian Federation	3134	4063	3783	10980	
48	Saudi Arabia	n/a	n/a	2929	2929	
49	Scotland	1049	959	n/a	2008	yes
50	Singapore	5420	4963	5532	15915	
51	Slovak Republic	n/a	4449	4570	9019	yes
52	Slovenia	2275	n/a	n/a	2275	yes
53	South Africa	n/a	4028	1284	5312	
54	Spain	n/a	1843	2877	4720	yes
55	Sweden	n/a	2491	2062	4553	yes
56	Trinidad and Tobago	n/a	n/a	2348	2348	
	Total observations	44590	104350	129159	278099	143917

Table 2: Descriptive statistics – PIRLS surveys 2001-2006-2011.

Variable	Obs	Mean	Std. Dev.	Min	Max
Individual students' characteristics					
PIRLS standardized test score	278099	523.38	86.94	66.44	775.36
Female	278099	0.51	0.5	0	1
Age (in months)	278099	123.81	7.56	76	165
Different language spoken at home	278099	0.28	0.45	0	1
Home education resources (high)	278099	0.12	0.33	0	1
Home education resources (medium)	278099	0.81	0.39	0	1
Home education resources (low)	278099	0.07	0.25	0	1
Parental education: tertiary	278099	0.3	0.46	0	1
Parental education: post-secondary	278099	0.21	0.41	0	1
Parental education: upper secondary	278099	0.31	0.46	0	1
Parental education: lower secondary	278099	0.12	0.32	0	1
Parental education: primary or less	278099	0.07	0.25	0	1
Schools' characteristics (weighted by students)					
Total number of students	278099	578.95	493.72	7	10916
Students in the 4th grade	278099	82.94	77.98	1	968
Urban	278099	0.5	0.5	0	1
Share of disadvantaged students: 0-10%	278099	0.38	0.49	0	1
Share of disadvantaged students: 10-25%	278099	0.28	0.45	0	1
Share of disadvantaged students: 25-50%	278099	0.18	0.38	0	1
Share of disadvantaged students: above 50%	278099	0.16	0.36	0	1
Average tenure of teachers	278099	17.83	9.77	0	51
Presence of a library	278099	0.89	0.31	0	1
Computers for 4th grade students	278099	0.36	0.61	0	48
Average class students' characteristics (weighted by students)					
Share of females	278099	0.5	0.18	0	1
Class size	278099	25.75	7.69	1	119[12]
Average age (in months)	278099	123.91	5.7	96	158
Share of students speaking a different language at home	278099	0.29	0.27	0	1
Share of students with low househ. education resources	278099	0.12	0.15	0	1
Share of students with medium househ. educ. resources	278099	0.81	0.18	0	1
Share of students with high househ. education resources	278099	0.07	0.15	0	1
Share of students with parental education: tertiary	278099	0.29	0.24	0	1
Share of students with parental education: post-secondary	278099	0.21	0.18	0	1
Share of students with parental education: upper secondary	278099	0.31	0.21	0	1
Share of students with parental education: lower secondary	278099	0.12	0.15	0	1
Share of students with parental education: primary or less	278099	0.07	0.15	0	1
Teachers' characteristics (weighted by students)					
Female teacher	278099	0.85	0.36	0	1
Age group: under 25	278099	0.03	0.16	0	1
Age group: 25-29	278099	0.11	0.31	0	1
Age group: 30-39	278099	0.29	0.46	0	1
Age group: 40-49	278099	0.31	0.46	0	1
Age group: 50-59	278099	0.23	0.42	0	1
Age group: 60 or more	278099	0.03	0.18	0	1
Graduated teacher	278099	0.91	0.28	0	1
Tenure (years)	278099	17.81	10.66	0	51

[12] Even if this value can seem unrealistic, the 99th percentile is 47 and therefore it does not affect any of the results of regression analysis.

Table 3: Institutional features

	Obs.	Mean	Std. Dev. (overall)	Std. Dev. (between)	Std. Dev. (within)	Min	Max
Any training before teaching	112	.88	.32	.25	.22	0	1
Examination required	101	.69	.46	.43	.22	0	1
Probationary teacher period	110	.51	.50	.46	.25	0	1
Length of probationary teacher period (months)	110	8.30	10.41	9.07	5.24	0	48
Mentoring or induction program	112	.33	.47	.39	.31	0	1
License or certification	112	.79	.41	.35	.27	0	1
Certification by Ministry of Education	103	.35	.48	.44	.20	0	1
Certification by National License board	103	.11	.31	.25	.18	0	1
Certification by University/College	103	.53	.50	.45	.24	0	1
Certification by Teacher organization/Union	103	.03	.17	.12	.11	0	1
Specific preparation on reading curricula	101	.84	.37	.35	.16	0	1
Number of certifications needed	103	.99	.75	.66	.43	0	3
Selectivity index	112	.67	.23	.21	.13	0	1
Pay index (salary/GDP per capita)	72	1.09	0.30	0.30	0.09	0.24	1.75

Table 4: Pair-wise correlation between institutional features

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Any training before teaching	1.00								
Examination required	0.11	1.00							
Probationary teacher period	0.14	-0.01	1.00						
Length of probationary teacher period (months)	0.08	-0.06	0.79*	1.00					
Mentoring or induction program	0.00	0.08	0.34*	0.22	1.00				
License or certification	0.11	0.34*	0.16	0.15	0.10	1.00			
Specific preparation on reading curricula	0.21	-0.10	0.28	0.25*	0.18	0.11	1.00		
Number of certifications needed	0.13	0.27*	0.24*	0.19	0.13	0.73*	0.08	1.00	
Selectivity index	0.43*	0.45*	0.65*	0.49*	0.57*	0.55*	0.49*	0.52*	1.00
Pay index (salary/GDP per capita)	0.15	-0.07	0.51*	0.48*	0.18	-0.12	0.14	-0.11	0.31

* Statistically significant at 1% level.

Table 5: Descriptive statistics of the reforms on teachers' labour market conditions (1947 – 2016).

	Obs.	Mean	Std. Dev. (overall)	Std. Dev. (between)	Std. Dev. (within)	Min	Max
Recruitment process (normalized)	3920	0.20	0.36	0.22	0.29	0	1
Working conditions (normalized)	3920	0.23	0.40	0.18	0.36	0	1
Salary conditions (normalized)	3920	0.24	0.42	0.25	0.34	0	1
Retirement (normalized)	3920	0.16	0.35	0.18	0.31	0	1

Table 6: Pair-wise correlation between reforms on teacher labour market conditions (1947–2016).

	Recruitment process (normalized)	Working conditions (normalized)	Salary conditions (normalized)	Retirement (normalized)
Recruitment process (normalized)	1.00			
Working conditions (normalized)	0.51*	1.00		
Salary conditions (normalized)	0.40*	0.49*	1.00	
Retirement (normalized)	0.39*	0.52*	0.37*	1.00

* Statistically significant at 1% level . 3920 observations.

Table 7. Pupils' score determinants

Sample:	Main sample			Reduce OECD sample		
Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)
Pupils' characteristics						
Female	14.17***	13.04***	13.46***	11.43***	10.92***	10.76***
	[0.280]	[0.278]	[0.248]	[0.358]	[0.332]	[0.323]
Age in months	-0.31***	-0.18***	-0.13***	-0.43***	-0.29***	-0.36***
	[0.027]	[0.027]	[0.025]	[0.037]	[0.034]	[0.036]
Different language spoken at home	-7.38***	-7.25***	-7.26***	-10.99***	-11.52***	-10.76***
	[0.338]	[0.341]	[0.322]	[0.459]	[0.428]	[0.429]
Index of home educational resources: Medium	-20.39***	-27.57***	-21.16***	-22.56***	-24.15***	-23.49***
	[0.437]	[0.453]	[0.421]	[0.555]	[0.538]	[0.520]
Index of home educational resources: Low	-32.12***	-43.87***	-32.65***	-42.62***	-49.32***	-44.92***
	[0.882]	[0.893]	[0.845]	[1.415]	[1.361]	[1.353]
Highest Parental Education: Tertiary or post	48.14***	42.00***	47.41***	48.75***	46.49***	47.08***
	[0.782]	[0.832]	[0.742]	[1.200]	[1.176]	[1.151]
Highest Parental Education: Post-secondary	35.71***	34.75***	35.82***	38.24***	37.14***	37.19***
	[0.767]	[0.802]	[0.723]	[1.209]	[1.170]	[1.155]
Highest Parental Education: Upper secondary	24.02***	23.08***	23.94***	27.32***	26.42***	25.90***
	[0.732]	[0.778]	[0.689]	[1.169]	[1.142]	[1.122]
Highest Parental Education: Lower secondary	8.72***	11.56***	9.25***	8.73***	10.77***	8.50***
	[0.725]	[0.751]	[0.682]	[1.189]	[1.152]	[1.134]
Teachers' characteristics						
Female	4.13***	5.90***	4.73***	3.48***	2.84***	2.42***
	[1.057]	[1.274]	[0.918]	[1.036]	[1.002]	[0.859]
Age group: Under 25	6.37*	20.68***	5.58**	-0.04	10.61***	-1.65
	[3.414]	[3.729]	[2.839]	[3.537]	[3.515]	[3.023]
Age group: 25-29	5.17*	15.63***	8.63***	-3.74	8.57***	-0.95
	[2.667]	[3.079]	[2.214]	[2.981]	[2.710]	[2.322]
Age group: 30-39	4.98**	9.66***	6.75***	-0.57	6.43***	-0.22
	[2.254]	[2.552]	[1.859]	[2.631]	[2.426]	[2.056]
Age group: 40-49	4.80**	11.65***	4.76***	-0.11	4.73**	-1.00
	[1.915]	[2.171]	[1.585]	[2.347]	[2.194]	[1.820]
Age group: 50-59	4.56***	11.87***	3.55**	1.41	5.29**	-0.39
	[1.716]	[1.975]	[1.423]	[2.134]	[2.062]	[1.642]
Tertiary education	7.11***	5.07***	4.04***	5.80**	-11.50***	1.19
	[2.044]	[1.622]	[1.567]	[2.835]	[1.843]	[1.869]
Tenure	0.10	0.18**	0.20***	-0.05	0.13	0.03
	[0.060]	[0.090]	[0.067]	[0.063]	[0.081]	[0.068]
Class' characteristics						
Female (%)	22.91***	7.54***	12.47***	14.66***	7.29***	5.79***
	[2.924]	[2.459]	[1.586]	[3.422]	[2.408]	[2.123]
Age in months (average)	-0.19	0.94***	2.33***	-0.39	1.66***	0.41**
	[0.230]	[0.087]	[0.131]	[0.258]	[0.085]	[0.189]
Different language spoken at home (%)	-10.75***	-7.69***	-11.10***	-18.83***	-29.29***	-19.06***
	[2.572]	[2.056]	[1.681]	[3.146]	[1.873]	[1.958]
Index of home educational resources: High (%)	46.22***	206.52***	47.73***	41.99***	122.10***	66.28***
	[7.166]	[7.041]	[5.723]	[10.186]	[8.122]	[7.722]
Index of home educational resources: Medium (%)	27.09***	81.23***	18.33***	31.30***	91.29***	44.13***
	[6.284]	[5.977]	[4.790]	[9.517]	[7.461]	[7.191]
Highest Parental Education: Tertiary or post (%)	96.46***	39.80***	95.53***	47.57***	34.06***	33.58***
	[6.756]	[6.435]	[4.896]	[7.593]	[6.445]	[5.731]
Highest Parental Education: Post-secondary (%)	78.03***	97.56***	89.21***	35.46***	36.47***	29.00***
	[6.706]	[6.167]	[4.914]	[7.670]	[6.307]	[5.733]
Highest Parental Education: Upper secondary (%)	53.48***	70.26***	61.52***	19.15***	21.80***	6.89
	[6.475]	[6.076]	[4.648]	[7.430]	[6.193]	[5.643]
Highest Parental Education: Lower secondary (%)	44.24***	107.18***	58.72***	13.33*	52.85***	15.99***
	[6.584]	[6.010]	[4.804]	[7.824]	[6.563]	[5.998]
Number of students	0.93***	-0.64***	-0.29***	0.62***	-0.11	0.03
	[0.123]	[0.086]	[0.066]	[0.147]	[0.085]	[0.064]
Schools' characteristics						
Total number of students		-0.00**	0.00**		-0.01***	0.00
		[0.002]	[0.001]		[0.002]	[0.002]
Students in the 4th grade		0.12***	0.02**		0.13***	0.00
		[0.012]	[0.008]		[0.017]	[0.016]
Urban		4.69***	1.62***		2.57***	0.22
		[0.872]	[0.596]		[0.689]	[0.586]

Share of disadvantaged students: 0-10%		22.09***	11.69***		13.91***	14.50***
		[1.522]	[1.089]		[1.419]	[1.259]
Share of disadvantaged students: 10-25%		20.63***	8.89***		12.77***	11.37***
		[1.521]	[1.067]		[1.409]	[1.245]
Share of disadvantaged students: 25-50%		11.90***	5.93***		9.03***	7.67***
		[1.656]	[1.107]		[1.495]	[1.306]
Average tenure of teachers		0.11	0.10		0.07	0.09
		[0.082]	[0.060]		[0.073]	[0.060]
Presence of a library		12.57***	7.63***		-3.03***	1.03
		[1.614]	[1.142]		[1.024]	[0.903]
Computers for 4th grade students		-1.25*	-0.88*		1.27*	-1.67**
		[0.641]	[0.463]		[0.698]	[0.704]
Constant	425.99***	219.78***	48.90***	536.14***	218.48***	423.36***
	[32.360]	[12.997]	[18.353]	[36.090]	[12.806]	[25.962]
School fixed effects	Yes	No	No	Yes	No	No
Country fixed effects	No	No	Yes	No	No	Yes
Wave fixed effects	No	Yes	Yes	No	Yes	Yes
Observations	278,099	278,099	278,099	143,917	143,917	143,917
Number of countries	56	56	56	29	29	29
R-squared	0.612	0.327	0.482	0.415	0.244	0.298

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets.

Table 8. Institutional features and pupils' test scores

Sample:	Main sample		Reduced OECD sample				
Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any training before teaching (dummy)	2.70*		1.81	9.23***			
	[1.476]		[2.098]	[2.193]			
Passing an Examination (dummy)	7.18***		6.64***	5.43***			
	[0.981]		[0.921]	[0.957]			
Length of probationary period	0.73***		0.74***	0.93***			
	[0.038]		[0.038]	[0.036]			
Probationary period (dummy)	17.29***		10.80***	16.39***			
	[0.912]		[0.909]	[0.851]			
Completion of a Mentoring or	0.85		4.54***	6.11***			
Induction Program (dummy)	[0.885]		[0.827]	[0.837]			
Specific preparation in how to	29.33***		16.85***	19.48***			
teach reading (dummy)	[1.180]		[1.171]	[1.193]			
Number of certifications needed	7.71***		-10.12***	-9.59***			
	[0.683]		[0.592]	[0.615]			
Process to license or certify	7.90***		-4.26***	-6.20***			
	[0.998]		[0.835]	[0.859]			
Overall index of selectivity		35.56***			40.56***		105.62***
		[2.106]			[2.202]		[9.599]
Pay index (salary/GDP per capita)			included			26.52***	68.02***
						[1.334]	[5.948]
Index of selectivity × pay index							-68.72***
							[8.326]
Number of countries	56	56	29	29	29	29	29
Observations	278,099	278,099	143,917	143,917	143,917	143,917	143,917

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions include controls for **pupils** (gender, age in months, a dummy for different language spoken at home, index of home educational resources, parental education), **teacher** (gender, age in 10-year group, educational certificate), **class** (share of females, average age in months, share of pupils speaking a different language at home, share of pupils with high home educational resources, class size) and **school characteristics** (total number of students, total number of students in the 4th grade, a dummy for schools in urban areas, share of disadvantaged pupils (in 4th grade), average tenure within the school, a dummy for the presence of a library, number of computers for pupils in the 4th grade) as listed in Table 7. Wave fixed effects are included. Columns 1, 3 and 4 report results for **distinct regressions where each institution is separately included by itself** and added to a regressions identical to Table 7, col.(2) and col.(5). R-squared (not reported for each cell) are similar to those in Table 7.

Table 9. Reforms on teachers' career and pupils' test scores.

Sample	Main	Reduced OECD		Main	Reduced OECD	
Variables included		All			One by one	
Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)
Recruitment reforms	4.51***	2.77***	2.62**	3.19***	2.21**	2.08**
	[1.174]	[1.069]	[1.066]	[1.211]	[1.055]	[1.053]
Working conditions reforms	-2.89**	-1.97	-1.85	-2.53**	-1.88	-1.82
	[1.233]	[1.211]	[1.211]	[1.145]	[1.184]	[1.184]
Salary reforms	-1.01	0.35	0.21	-1.50	-0.15	-0.24
	[1.084]	[1.043]	[1.047]	[1.018]	[1.007]	[1.011]
Retirement reforms	-1.46	-2.53**	-2.43**	-1.52	-2.52**	-2.45**
	[1.338]	[1.209]	[1.210]	[1.329]	[1.204]	[1.204]
Pay index		13.10*			included	
		[6.859]				
Number of countries	56	29	29	56	29	29
Observations	278,099	143,917	143,917	278,099	143,917	143,917

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions include pupils, teacher, class and school characteristics as listed in Table 7. Country and wave fixed effects are also included. Coefficients for pay index are always positive and significant at 10% level.

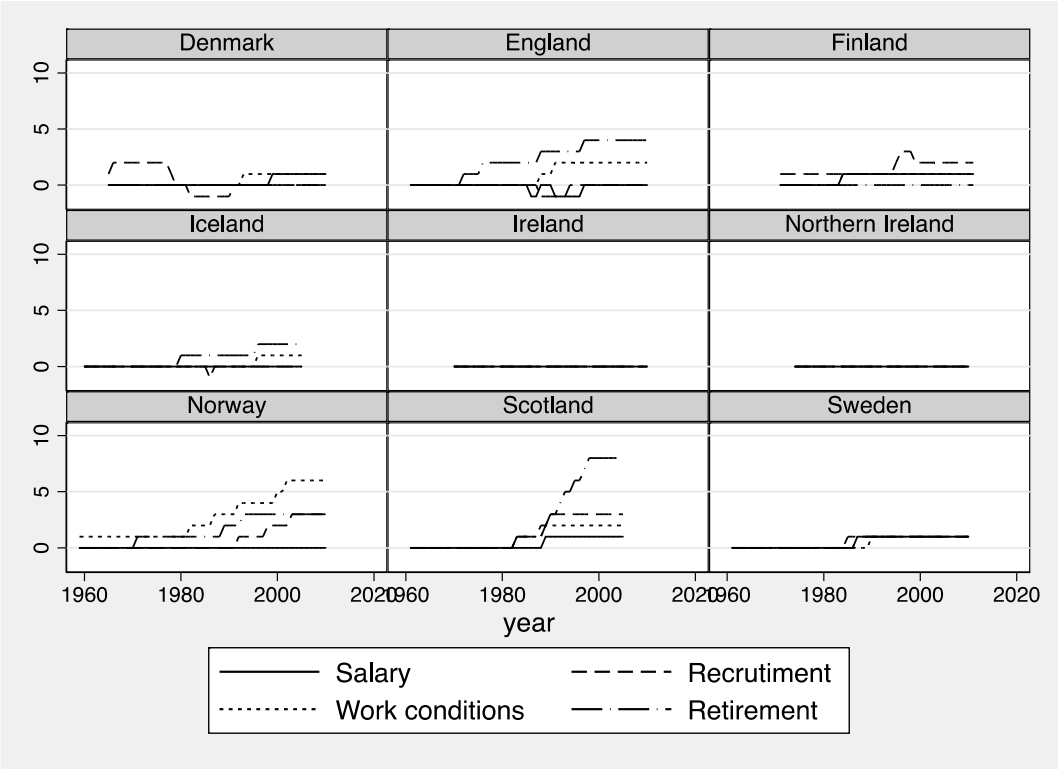
Table 10. Reforms on teachers' career and pupils' test scores.

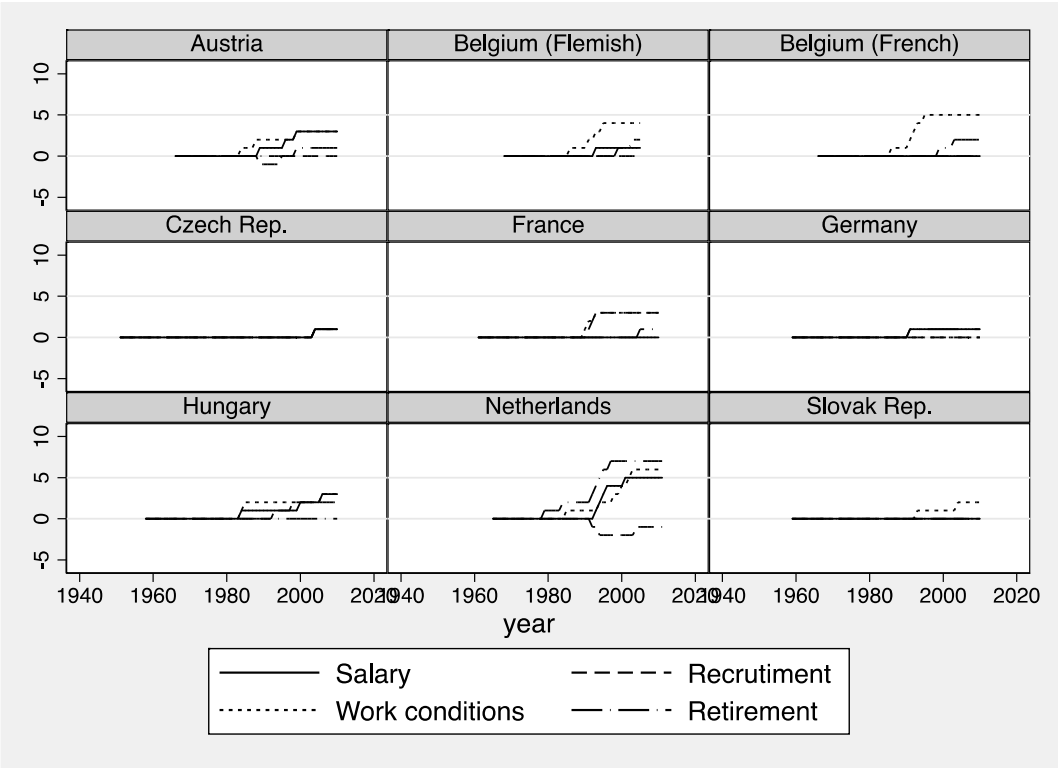
Sample	Main	Reduced OECD		Main	Reduced OECD	
Variables included		All			One by one	
Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)
Recruitment reforms	4.17***	4.35***	4.43***	-0.34	3.38***	3.50***
	[1.035]	[0.825]	[0.830]	[1.098]	[0.808]	[0.819]
Working conditions reforms	-13.85***	-10.08***	-13.41***	-11.48***	-7.74***	-11.21***
	[1.116]	[1.007]	[1.022]	[1.111]	[0.949]	[0.971]
Salary reforms	6.78***	8.87***	13.32***	2.31**	7.78***	12.58***
	[1.019]	[0.857]	[0.842]	[1.043]	[0.845]	[0.836]
Retirement reforms	-8.73***	-2.08**	-2.16**	-10.37***	-3.13***	-3.93***
	[1.263]	[0.990]	[1.006]	[1.306]	[0.949]	[0.982]
Pay index		19.75***			included	
		[1.369]				
Number of countries	56	29	29	56	29	29
Observations	278,099	143,917	143,917	278,099	143,917	143,917

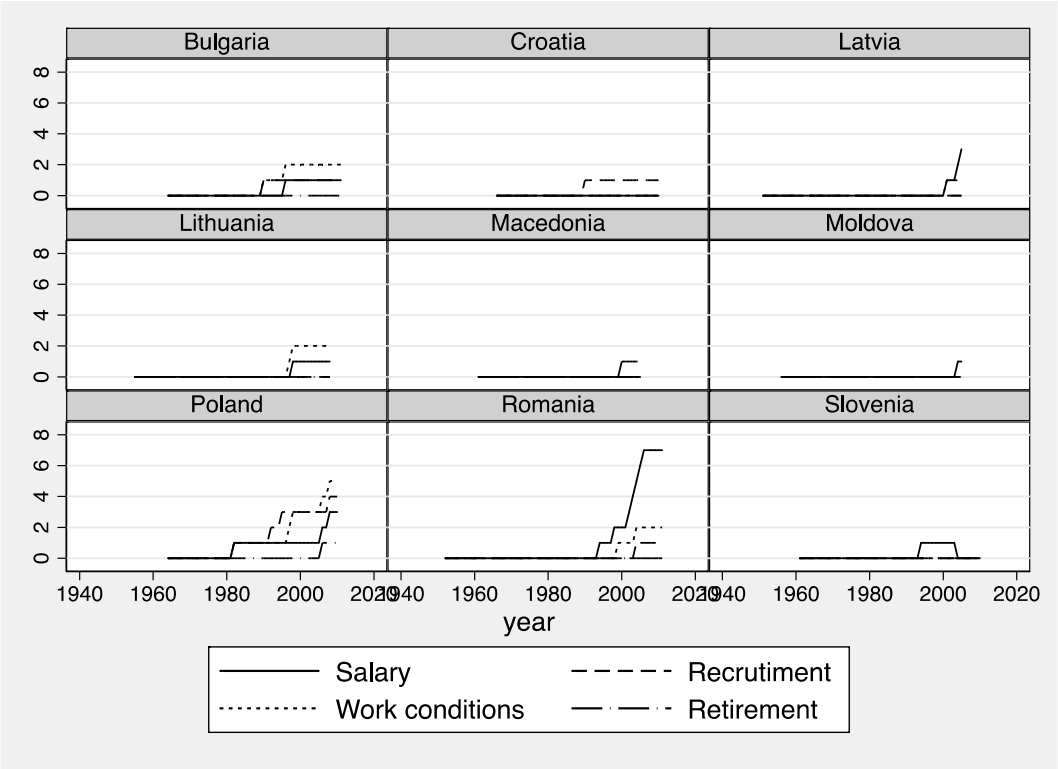
*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions include pupils, teacher, class and school characteristics as listed in Table 7. Wave fixed effects are included. Coefficients for pay index are always positive and significant at 1% level.

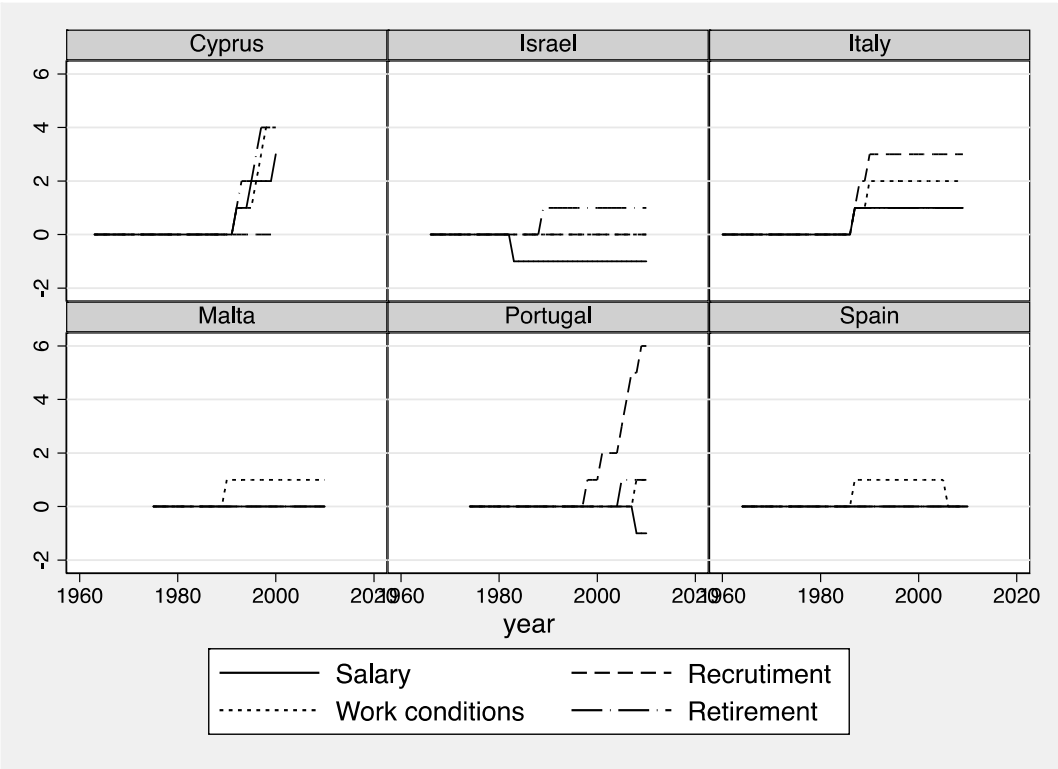
FIGURES

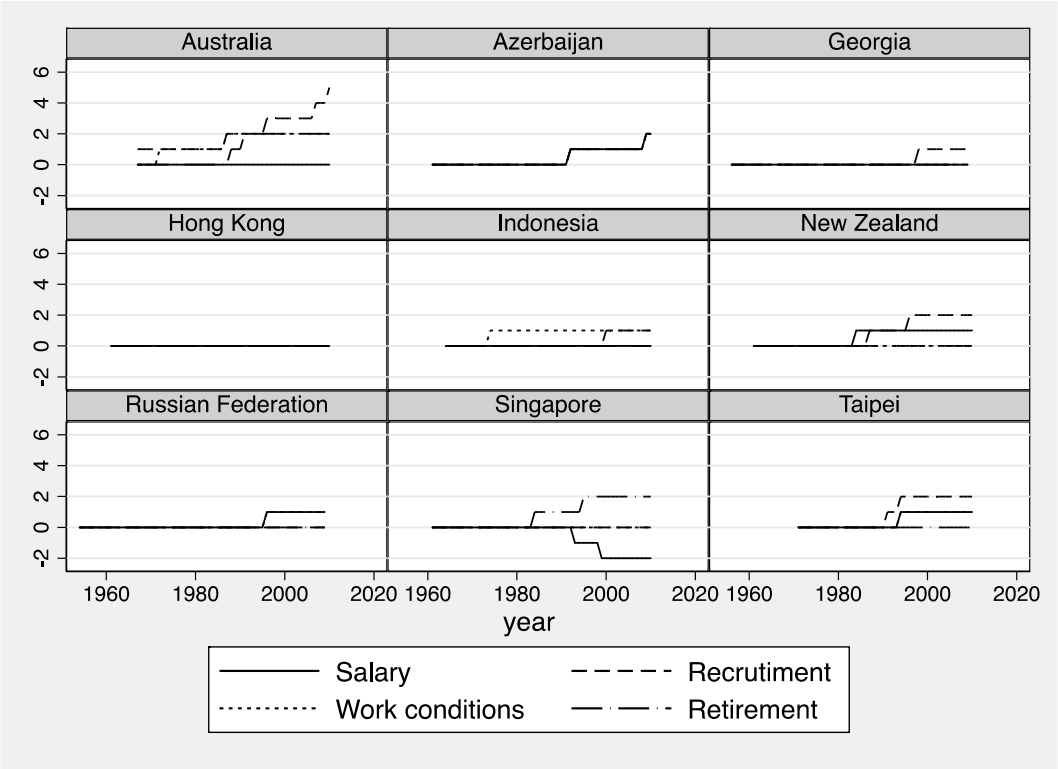
Figure 1. Reforming activity in teacher policies by country and year of intervention

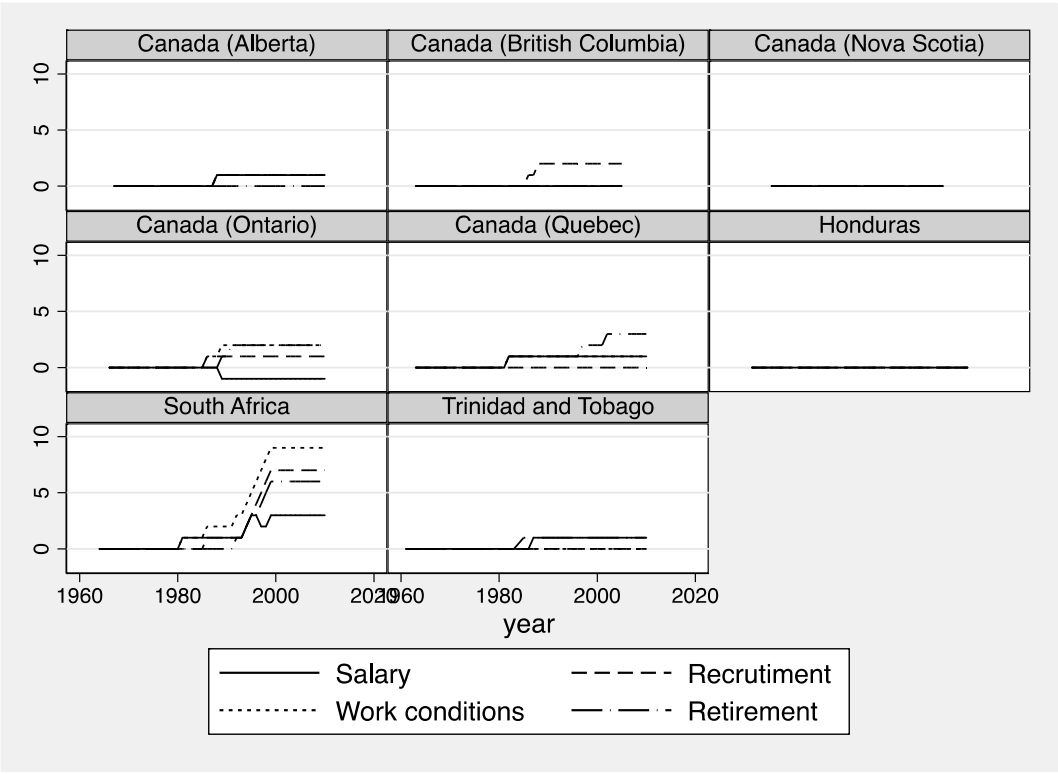


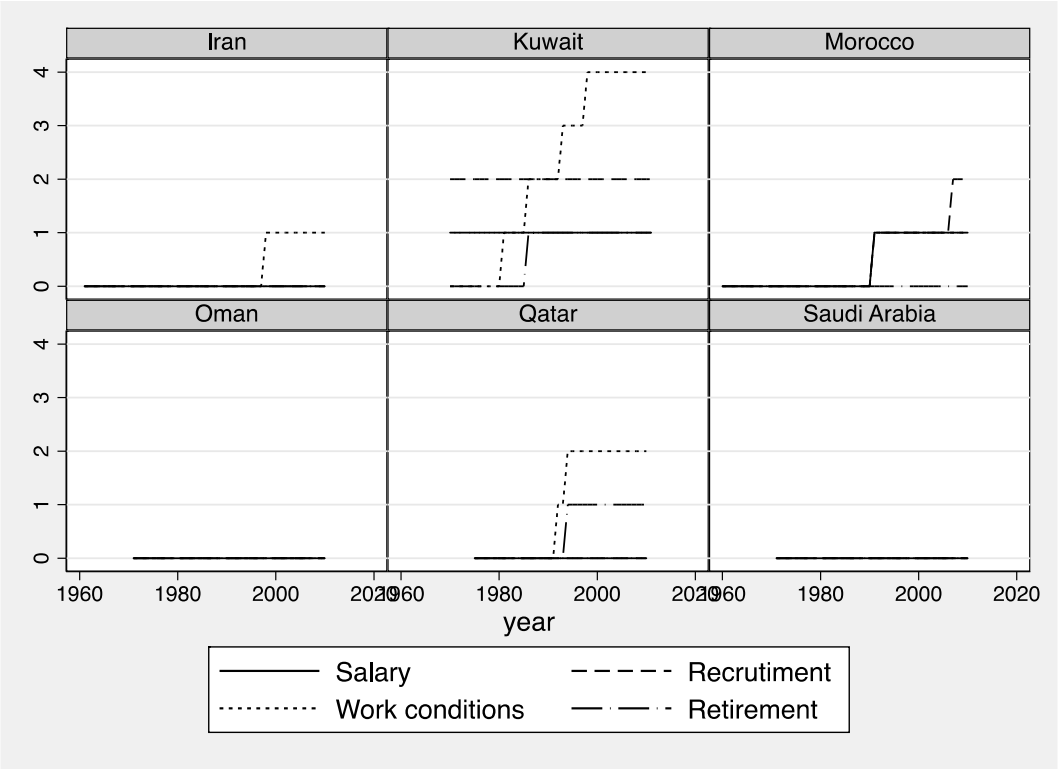



Bulgaria
Croatia
Latvia
Lithuania
Macedonia
Moldova
Poland
Romania
Slovenia
0 2 4 6 8
1940 1960 1980 2000 2020
year
Salary
Recrutiment
Work conditions
Retirement


Cyprus
Israel
Italy
Malta
Portugal
Spain
-2 0 2 4 6
1960 1980 2000 2020
year
Salary
Recrutiment
Work conditions
Retirement


Australia
Azerbaijan
Georgia
Hong Kong
Indonesia
New Zealand
Russian Federation
Singapore
Taipei
-2 0 2 4 6
1960 1980 2000 2020
year
Salary
Recrutiment
Work conditions
Retirement


Canada (Alberta)
Canada (British Columbia)
Canada (Nova Scotia)
Canada (Ontario)
Canada (Quebec)
Honduras
South Africa
Trinidad and Tobago
0 5 10
1960 1980 2000 2020
year
Salary
Recrutiment
Work conditions
Retirement


Iran
Kuwait
Morocco
Oman
Qatar
Saudi Arabia
0
1
2
3
4
1960
1980
2000
2020
year
Salary
Work conditions
Recrutiment
Retirement

APPENDIX

A.1 Further analysis and robustness checks

Table A.1.1. Institutional features and pupils' test scores.
Heterogeneous effects by teacher educational level

Sample size:	Main sample			Reduced OECD sample		
Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)
Teachers' education:	Graduate	Non graduate	Difference	Graduate	Non graduate	Difference
Any training before teaching (dummy)	7.23***	-17.14***	24.37***	20.95***	-10.27***	31.21***
	1.602	3.320	3.633	2.699	3.268	4.166
Passing an Examination (dummy)	7.20***	6.99*	0.20	3.89***	66.02***	-62.13***
	1.002	3.961	4.052	0.906	9.274	9.324
Length of probationary period	0.82***	0.27**	0.56***	0.91***	1.04***	-0.12
	0.0413	0.105	0.114	0.0373	0.107	0.111
Probationary period (dummy)	18.65***	-6.47	25.12***	15.22***	48.55***	-33.33***
	0.921	4.681	4.777	0.834	5.095	5.116
Completion of a Mentoring or	3.33***	-21.74***	25.07***	5.83***	9.76***	-3.93
Induction Program (dummy)	0.925	2.903	3.046	0.864	3.173	3.279
Specific preparation in how to	32.38***	-7.28*	39.66***	21.94***	2.66	19.28***
teach reading (dummy)	1.218	3.854	4.048	1.248	3.561	3.778
Process to license or certify	8.49***	-2.67	11.16***	-9.08***	-23.56***	14.47***
	0.704	2.456	2.547	0.617	4.246	4.288
Number of certifications needed	8.75***	-17.85***	26.60***	-6.23***	-3.87	-2.36
	1.004	6.462	6.535	0.864	6.063	6.116
Overall index of selectivity	41.88***	-43.22***	85.11***	38.78***	60.83***	-22.05*
	2.159	8.104	8.436	2.228	11.44	11.77
Pay index				24.94***	48.46***	-23.52***
				1.326	7.184	7.324
Number of countries	56	56	56	29	29	29
Observations	278,099	278,099	278,099	143,917	143,917	143,917

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions include pupils, teacher, class and school characteristics as listed in Table 7. Wave fixed effects are included. The table includes regressions analogous to col.(1) and (3) of Table 8, where interacted terms are added. In particular, col.(1)-(3) report the marginal effect of each institution for graduate teachers (col.1), non-graduate teachers (col.2) and the difference (col.3), analogous to the models in Table 8, col.(1). Col.(4)-(6) report the same effects on the restricted sample for which information on pay are available (Table 8, col.3).

Table A.1.2. Reforms on teachers' career and pupils' test scores.
Heterogeneous effects by teacher educational level

Sample size:	Main sample			Reduced OECD sample			Reduced OECD sample		
Teachers' education	Graduate	Non graduate	Difference	Graduate	Non graduate	Difference	Graduate	Non graduate	Difference
Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Recruitment reforms	-1.08	12.27**	-13.35***	2.85***	11.84***	-8.99**	3.27***	7.33*	-4.064
	1.108	4.919	4.984	0.812	4.144	4.203	0.821	4.365	4.426
Salary reforms	2.36**	1.78	0.58	7.40***	13.22***	-5.82*	12.51***	13.66***	-1.144
	1.060	4.022	4.114	0.859	3.389	3.454	0.839	3.449	3.462
Working conditions reforms	-11.97***	-4.20	-7.77**	-7.50***	-12.51***	5.004	-10.69***	-20.63***	9.94**
	1.126	4.277	4.331	0.944	4.538	4.526	0.955	4.819	4.797
Retirement reforms	-11.73***	12.00**	-23.73***	-4.07***	8.96**	-13.03***	-4.80***	7.25**	-12.05***
	1.329	4.670	4.761	0.962	3.527	3.586	0.999	3.626	3.699
Pay index	No	No	No	Yes	Yes	Yes	No	No	No
Number of countries	56	56	56	29	29	29	29	29	29
Observations	278,099	278,099	278,099	143,917	143,917	143,917	143,917	143,917	143,917

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions include pupils, teacher, class and school characteristics as listed in Table 7. Wave fixed effects are included. The Table includes regressions analogous to col.(3) of Table 10, where interacted terms are added. In particular, col.(1)-(3) report the marginal effect of each institution for graduate teachers (col.1), non-graduate teachers (col.2) and the difference (col.3). Similarly for columns (4)-(6) and (7)-(9).

Table A.1.3. Reforms during the teacher career

Sample	Main	Reduced OECD		Main	Reduced OECD	
Variables included		All			One by one	
Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)
Recruitment reforms	0.95***	0.14	0.18	-2.65***	1.15***	0.84***
	[0.290]	[0.219]	[0.228]	[0.336]	[0.202]	[0.209]
Working conditions reforms	-6.89***	-2.08***	-3.92***	-6.37***	0.54*	-0.77***
	[0.384]	[0.353]	[0.338]	[0.380]	[0.293]	[0.292]
Salary reforms	2.46***	6.70***	7.79***	-1.10***	5.39***	5.69***
	[0.303]	[0.421]	[0.417]	[0.308]	[0.328]	[0.328]
Retirement reforms	-2.36***	-0.18	0.67***	-5.18***	0.48**	1.03***
	[0.372]	[0.264]	[0.254]	[0.439]	[0.244]	[0.236]
Pay index		23.69***			Included	
		[1.414]				
Number of countries	56	29	29	56	29	29
Observations	278,099	143,917	143,917	278,099	143,917	143,917

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions include pupils, teacher, class and school characteristics as listed in Table 7. Wave fixed effects are included. Coefficients for pay index are always significant at 10% level.

Table A.1.4. Institutional features and pupils' test scores by teachers' tenure.

Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)
Tenure:	0-5	0-10	0-20	0-30	All
Overall index of selectivity	56.60***	52.14***	45.30***	38.30***	35.56***
	[5.362]	[3.691]	[2.698]	[2.296]	[2.106]
Observations	42,211	86,274	170,167	240,685	278,099

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions are the same as Table 8, column (2), with the sample restricted by teachers' tenure.

Table A.1.5. Institutional features and pupils' test scores by teachers' tenure.

Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)
Tenure:	0-5	0-10	0-20	0-30	All
Pay index	35.52***	31.74***	29.77***	27.20***	26.52***
	[3.725]	[2.517]	[1.802]	[1.488]	[1.334]
Observations	22,984	46,051	85,703	123,802	143,917

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions are the same as Table 8, column (6), with the sample restricted by teachers' tenure.

Table A.1.6. Reforms on teachers' career and pupils' test scores by teachers' tenure.

Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)
Tenure:	0-5	0-10	0-20	0-30	All
Variables included:	All	All	All	All	All
Recruitment reforms	-4.60*	-1.35	0.51	3.23***	4.17***
	[2.636]	[1.829]	[1.363]	[1.143]	[1.035]
Working conditions reforms	-11.91***	-13.24***	-12.05***	-12.91***	-13.85***
	[2.438]	[1.640]	[1.260]	[1.139]	[1.116]
Pay reforms	5.74**	6.54***	6.94***	6.34***	6.78***
	[2.370]	[1.639]	[1.190]	[1.057]	[1.019]
Retirement reforms	-6.25***	-9.41***	-10.00***	-8.81***	-8.73***
	[2.298]	[1.591]	[1.335]	[1.279]	[1.263]
Salary index	No	No	No	No	No
Observations	42,211	86,274	170,167	240,685	278,099

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions are the same as Table 10, column (1), with the sample restricted by teachers' tenure.

Table A.1.7. Institutional features and pupils' test scores - Falsification test

Sample:	Main sample		Reduced OECD sample				
Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any training before teaching (dummy)	0.96**		1.11	-0.10			
	[0.438]		[0.879]	[0.515]			
Passing an Examination (dummy)	7.18***		6.63***	5.43***			
	[0.981]		[1.051]	[0.957]			
Length of probationary period	-0.00		0.01	0.01			
	[0.014]		[0.022]	[0.017]			
Probationary period (dummy)	0.17		0.41	0.26			
	[0.290]		[0.479]	[0.340]			
Completion of a Mentoring or	0.11		0.70	0.15			
Induction Program (dummy)	[0.305]		[0.456]	[0.357]			
Specific preparation in how to	0.60		0.33	0.50			
teach reading (dummy)	[0.452]		[0.658]	[0.533]			
Number of certifications needed	0.01		0.24	0.05			
	[0.205]		[0.334]	[0.235]			
Process to license or certify	-0.03		-0.09	0.03			
	[0.326]		[0.565]	[0.381]			
Overall index of selectivity		0.81			0.53		1.43
		[0.640]			[0.744]		[4.543]
Pay index			included			-1.06*	-0.70
						[0.542]	[2.779]
Index of selectivity × pay index							0.12
							[4.049]
Number of countries	56	56	29	29	29	29	29
Observations	244,011	244,011	126,709	126,709	126,709	126,709	126,709

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. All regressions are the same as Table 8, with the sample restricted by the random assignment procedure.

Table A.1.8. Reforms on teachers' career and pupils' test scores - Falsification test

Sample	Main	Reduced OECD		Main	Reduced OECD	
Variables included		All			All	
Dependent variable: Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)
Recruitment reforms	0.44	0.22	0.08	-0.40	-0.45	-0.40
	[0.465]	[0.635]	[0.546]	[0.491]	[0.630]	[0.547]
Working conditions reforms	0.45	0.51	0.00	0.43	-0.10	0.67
	[0.430]	[0.634]	[0.511]	[0.507]	[0.701]	[0.548]
Salary reforms	-0.35	0.67	-0.23	-0.08	-0.09	-0.07
	[0.397]	[0.563]	[0.467]	[0.418]	[0.575]	[0.478]
Retirement reforms	0.06	-1.22*	-0.07	-0.15	-0.09	-0.96*
	[0.405]	[0.626]	[0.475]	[0.477]	[0.673]	[0.526]
Pay index		-0.13			-0.48	
		[0.669]			[0.665]	
Number of countries	56	29	29	56	29	29
Observations	262,003	74,628	135,588	262,691	75,558	135,806

*** p<0.01, ** p<0.05, * p<0.1, robust standard errors clustered at school level in brackets. Regressions in blocks (1)-(3) and (4)-(6) are the same as Table 10, col. (1)-(3), the only difference being the random assignment procedure (among all observations in col.1-3 and across countries within the same cohort in col-4-6).

A.2 Institutional reforms

Table A.2.1 – List of reforms by country and year

Country	Year	(1)	(2)	(3)	(4)
Abu Dhabi	1947				
Argentina	1958	✓			
	1988	✓	✓	✓	✓
	1989		✓		
	1991		✓	✓	
	1993	✓			
	2007				✓
	2016	✓			
Australia	1964	✓			
	1972		✓		
	1987	✓	✓		
	1988			✓	
	1991			✓	
	1996	✓			
	2007	✓			
	2010	✓			
	2013	✓			
	2015	✓			
Austria	1984		✓		
	1988		✓		
	1989	✓			✓
	1995	✓			
	1996				✓
	1999		✓	✓	✓
Azerbaijan	1992	✓	✓	✓	✓
	2009	✓	✓	✓	✓
Belgium (Flemish)	1986		✓		
	1991		✓		
	1993		✓		✓
	1995		✓		
	1999			✓	
	2003			✓	
Belgium (French)	1986		✓		
	1991		✓		
	1992		✓		
	1993		✓		
	1995		✓		
	1999			✓	
	2003			✓	
Belize	1992		✓	✓	✓
	2010	✓			
Bulgaria	1990	✓	✓		
	1996		✓		✓
	2015	✓			✓
	2016	✓			
Canada, Alberta	1988	✓	✓		✓
Canada, British Columbia	1986	✓			
	1988	✓			
Canada, Ontario	1947				
	1986	✓	✓		
	1989		✓	✓	✓
Israel	1983				✓
	1989			✓	
Italy	1987	✓	✓	✓	✓
	1988	✓			
	1990	✓	✓		
Kuwait	1960	✓			✓
	1970	✓			
	1981		✓		
	1986		✓	✓	
	1993		✓		
	1998		✓		
Latvia	2001		✓		✓
	2004				✓
	2005				✓
	2016				✓
Lithuania	1997		✓		
	1998		✓		✓
Luxembourg	1991	✓	✓		
	1992	✓			
	1996		✓		
	2009	✓			
Macedonia	2000			✓	
	2008	✓			
	2016	✓			
Malta	1990		✓		
	2012		✓		
Moldova, Rep. Of	2004				✓
Morocco	1947				
	1991	✓	✓		✓
	2007	✓			
Netherlands	1979			✓	
	1984			✓	
	1985		✓		
	1992	✓		✓	
	1993			✓	✓
	1994	✓	✓	✓	✓
	1995			✓	✓
	1996				✓
	1997			✓	
	1998		✓		
	2000		✓		
	2001				✓
	2002		✓		
	2003		✓		
	2004	✓			
New Zealand	1984				✓
	1987	✓			
	1996	✓			
Northern Ireland	2014	✓			
Norway	1958		✓		
	1971			✓	

	1991			✓	
Canada, Quebec	1982		✓	✓	✓
	1997			✓	
	2002			✓	
Chinese Taipei	1991	✓			
	1994	✓	✓		✓
Colombia	1994		✓		✓
Croatia	1990	✓			
	2012	✓			
	2014	✓			
Cyprus	1992		✓	✓	✓
	1993			✓	
	1995				✓
	1996		✓	✓	
	1997		✓	✓	
	1998		✓		
	2000				✓
	2010			✓	
	2014	✓			
	2015		✓		
Czech Republic	2004	✓	✓	✓	✓
Denmark	1954	✓			
	1966	✓			
	1978	✓			
	1979	✓			
	1982	✓			
	1991	✓			
	1993		✓		
	1999				✓
Dubai	1947				
England	1972			✓	
	1976			✓	
	1986				✓
	1988	✓	✓	✓	✓
	1991		✓		✓
	1994	✓			
	1997			✓	✓
	2011	✓			
Finland	1971	✓			
	1984		✓		✓
	1995	✓			
	1996	✓			
	1999	✓			
France	1990		✓		
	1991	✓	✓		
	1992	✓			
	1993	✓	✓		
	2005			✓	
	2013	✓			
Georgia	1998	✓			
	2015	✓	✓		
Germany	1991			✓	✓
	2014		✓		
Greece	1984				✓
	1985	✓			
	1992	✓			
	1982		✓		
	1987		✓		
	1989			✓	
	1992	✓	✓		
	1993			✓	
	1998	✓			
	2000		✓		
	2002		✓		
Norway	2003	✓			
Oman	2012	✓	✓	✓	✓
Poland	1982	✓	✓		✓
	1992	✓			
	1995	✓			
	1997		✓		
	1998		✓		
	2006		✓	✓	✓
	2008	✓	✓		✓
	2012	✓			
Portugal	1998	✓			
	2001	✓			
	2005	✓		✓	
	2006	✓			
	2007	✓			
	2008		✓		✓
	2009	✓			
	2012	✓			
	2014	✓	✓		✓
	2015		✓		
Qatar	1992		✓		
	1994		✓	✓	
Romania	1994				✓
	1998				✓
	1999		✓		
	2002				✓
	2003				✓
	2004	✓	✓		✓
	2005				✓
Romania	2006				✓
Russian Federation	1996		✓		✓
	2012	✓			
Saudi Arabia	1947				
Scotland	1983	✓	✓	✓	
	1988			✓	
	1989	✓	✓		✓
	1990	✓		✓	
	1992			✓	
	1993			✓	
	1995			✓	
	1997			✓	
	1998			✓	
	2010			✓	
	2013			✓	
Singapore	1984			✓	
	1993				✓
	1995			✓	
	1999				✓

	2010	✓			Slovak Republic	1993		✓	
	2015			✓		2004		✓	
Honduras	2011	✓			Slovenia	1994			✓
Hong Kong	1947					2004			✓
Hungary	1984		✓	✓	South Africa	1981		✓	✓
	1985		✓			1986	✓	✓	
	1993	✓				1992		✓	✓
	1998	✓				1994	✓	✓	✓	✓
	2000			✓		1995	✓	✓	✓	✓
	2006		✓	✓		1996	✓	✓		
	2015	✓				1997	✓	✓	✓	✓
Iceland	1980			✓		1998	✓	✓	✓	
	1986	✓				1999	✓	✓	✓	✓
	1987	✓			Spain	1987		✓		
	1996		✓	✓		2006		✓		
	2015	✓			Sweden	1985	✓			
Indonesia	1974		✓			1987			✓	✓
	2000	✓				1990		✓		
	2014			✓	Trinidad and Tobago	1985			✓	
Iran	1998		✓			1987				✓
	2015	✓			Turkey	1999	✓			
Ireland	2011	✓				2001		✓		
	2012			✓		2012		✓		
					United States	1974	✓	✓	✓	✓
						1988	✓	✓	✓	✓

Note. Columns refer to: Recruitment process (1), Working conditions (2), Retirement (3), Pay (4).