

# Selecting or rewarding teachers?

## International evidence from primary schools<sup>♣</sup>

Michela Braga<sup>a</sup>, Daniele Checchi<sup>b</sup>, Christelle Garrouste<sup>c</sup>, Francesco Scervini<sup>d</sup>

<sup>a</sup>Bocconi University, Milan (Italy)

<sup>b</sup>University of Milan (Italy) and IZA

<sup>c</sup>Paris 12 Val de Marne University (France)

<sup>d</sup>University of Pavia (Italy)

*This version: January 25<sup>th</sup>, 2018*

**Abstract:** Using data from three waves of PIRLS, this paper examines the effect of teacher quality on pupils' literacy test scores at 4<sup>th</sup> grade by exploiting variations induced by reforms in teachers' selection and/or reward. We construct an original dataset of relevant reforms taking place at the national level over the last century and affecting the working conditions of primary schools teachers, matching them by the year when they entered the profession. After showing that teacher experience/age and qualification are significantly correlated with student competences, 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 reform introduction and entry into the profession. Results point to more selective recruitment process and, at a lesser extent, more generous wage conditions as effective ways to enhance pupils' performance.

**JEL code:** H52 (*Government Expenditures and Education*), I21 (*Analysis of Education*), I28 (*Government Policy*), J44 (*Professional Labour Markets – Occupational Licensing*)

**Keywords:** *Student achievements, PIRLS, Teacher recruitment process, Teacher pay.*

---

\*Paper presented at 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 helpful comments. All errors are our own responsibility.

## 1. Introduction

Increasing the quantity and the quality of education, in terms of years of schooling and students achievements, is a central goal for policy makers. Finding effective tools to achieve this goal is a hot issue in the academic debate. However, nowadays, especially in advanced economies, more than increasing the access to education, it is even more important to develop the right skills and to convey knowledge to students. Poor education quality penalizes pupils for life since it translates into worse life prospects in terms of lower lifetime earnings, higher difficulties in adapting to modern knowledge-based economies or higher unemployment probability. Since educational failure is costly both for the individual and the society, enhancing educational standards is a priority in the policy agenda of many countries.

It is not simple to identify which factors contribute in shaping a good schooling environment. The quality of the educational systems does not depend on a single factor but on a variety of elements, among which curriculum, learning environment, organization of class activities and teaching techniques. Teachers with different competencies can clearly influence these elements and, therefore, represent a key input for the educational process. Teachers' quality is an essential element in determining schools' quality. Basing on administrative data, a growing body of empirical literature shows that teachers' quality matters and it is the most important school input predicting students' as well as adults' learning gains (Hanushek, 2011; Chetty et al. 2014a and 2014b), but 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 can only focus on observable characteristics (i.e. credentials, experience accumulated on-the-job, formal and informal training), which tend to be weakly correlated with teachers' contribution to pupils' achievement. As a consequence, unobservable features, such as ability or effort, emerge as significantly correlated with teacher quality.<sup>1</sup>

Since the teaching staff costs represent, on average, two thirds of 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 best teachers. Two possible drivers to attain this goal are represented by selection or remuneration practices implemented by school principals. However, since both selection and pay policies could be correlated with unobservable characteristics, evaluating which one is the most effective tool is a very difficult task. Indeed, the presence/absence of monetary incentives in the profession can induce positive/negative self-selection of individuals. On top of that, it is not clear what is the relationship between the two policy instruments. Namely, it is hard to appraise whether selection and remuneration are substitute or complement tools to enhance teacher's quality. On the one hand, selectivity

---

<sup>1</sup> Goldhaber et al. (1999) estimate that the share of teacher effects due to unobservable characteristics can be as high as 97%.

should attract better-endowed teachers, both in terms of observables and unobservables. On the other hand, merit-pay wage policies should attract people expecting to benefit most from such a scheme, but it is hard to anticipate whether they are abler or greedier than average. As a result, it is almost impossible to predict the overall effect on students' achievement, since "selection" and "incentive" effects may work in opposite directions.

This paper contributes to the debate on the quality of educational systems in different directions. First, differently from most previous literature focused on secondary schools, we consider the role exerted by teachers on cognitive skills formation in primary schools, the first and essential stage in competence acquisition. Second, a newly assembled dataset on the institutional features relevant for primary school teachers over the last century allows exploiting the exogenous variation in these features to identify in a causal sense the impact of teacher's quality on pupil performance. Third, as far as we know, this is the first paper exploring the issue of substitutability and complementarity among alternative policy measures enhancing school quality. In addition, while most scholars focus on a single country and on a single reform, we conduct our analysis in a comparative perspective and we exploit results from an international comparable standardized assessment that combines information on pupils' performance with information on their teachers. Obviously, the chosen approach has to face the 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 including 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 together with information on the teaching framework for primary teachers in the year of the pupils' assessment. Then, we explore the relationship between teacher working conditions and pupil achievement. More specifically we proxy individual teachers' quality with group averages, where 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 scores of fourth graders. Our identification strategy relies on temporal and geographical variations in the institutional arrangements, controlling for individual, class, school and teacher characteristics. This strategy is not new in the literature, but so far it has focused only on one type of reform at a time (Braga et al. 2013), while we take into account the fact that a single cohort of students may be affected by teachers hit by different types of reforms identified according to their age.

The econometric analysis indicates that policies can significantly improve pupils' performance, via a possible change in teachers' quality due to the introduction of more selective recruitment process and/or more generous pay conditions. Our results have policy implications both in the short and in the long run. In the short run, being able to attract and retain best teachers is a cost-effective strategy. But it can have also positive spillovers in the long run. Indeed, basic and essential reading, writing and numerical

abilities are formed and developed especially in primary education. These three abilities are fundamental for intellectual capabilities in future life. Without them, any investment in skill formation in subsequent stages of a student educational process would be more costly and less effective. Any significant human capital accumulation requires solid foundations built in early stages of the learning process.

The rest of the paper is organized 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. 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.

## **2. Literature Review: the state of the art**

This paper is linked to two main strands of the literature: the one analysing practices to enhance the quality of educational systems and the other concerning the link between teachers' features and pupils' performance.

The literature on Human Resource management indicates three main goals for any employer with respect to manpower: recruiting (which implies attracting and selecting), motivating, and retaining the employees in order to achieve the highest level of productivity (Lazear and Gibbs, 2007). Since the ability to teach is a job-specific human capital, retaining teachers into the profession 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 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 with 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. have 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 becoming professionals. Nagler, Piopiunik and West (2015) obtain analogous results for more recent years by exploiting business cycle conditions at career start 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 and find a negative relationship between teacher shortages and regional unemployment rates over the period 1981-2002. They explain this effect by the centralized and rigid pay system in the public sector that tend to reduce labour supply and lead to shortages of qualified personnel. Hence, according to these authors, the teaching profession would remain a residual one due to 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 for 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 to other highly qualified professionals. There is evidence that teacher wage premium and wage dispersion have little effects on the quality mix of applicants to teaching. For instance, Ortega (2010) shows that, in the case of Venezuela, most students' preference for teaching is unresponsive to wage levels relative to other occupations and to wage growth prospects within teaching. If wages may 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 higher ability individuals into teaching and permitting scope for quicker salary advancement has a positive effect on pupil outcomes.

Thus, given the difficulty of ensuring that best candidates enter teacher training, it becomes even more crucial to ensure the highest quality in teacher training programs. Unfortunately, the literature does not report any consistent relationship between the level of credentials of teachers and corresponding students' 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 teachers' test scores and average students' achievement scores. The study by Andersson et al. (2011) on the effect of teacher certification on Swedish students' grade point average (GPA) shows that a one percentage point increase in the share of non-certified 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 American schools: they do not find any evidence that teacher' pre-service (undergraduate) training or college entrance exam scores being 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 schools shows that, on average, the initial certification status of a teacher has small impacts on students' test performance. Along the same line of research, the study by Buddin and Zamarro (2009) on teacher qualifications and student achievement in Los Angeles primary schools reveals that nor the teacher licensure<sup>2</sup> test scores nor the possession of an advanced degree are related to student achievement. While selecting skilled people into teaching is a key step towards more effective schools, academic research suggests that it can be improved upon by appropriate reward policies acting as a motivating device. In fact, a common concern of policy makers refers to the retention of best teachers into the profession. Indeed, the cost of a high turnover of high-quality teachers goes beyond the loss in students' performance. As reported by Watlington et al. (2010), when high-quality teachers leave the classroom, there is a significant negative effect on both the students and the school performance.

---

<sup>2</sup> "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).

If good teachers are to be retained in teaching and supported in doing their work at best, they should have a workplace that promotes their efforts in a variety of ways (Moor Johnson, 2006). Since the 1980s, U.S. and U.K. have passed measures to implement performance-based incentives, i.e. monetary benefits to teachers and/or school principals, considered the best in terms of existing levels and/or in terms of improvement in the level of student achievements (Holanda et al. 2008; Hanushek and Rivkin, 2006; Schächter and Thum, 2004). However, these policy measures have proven to have contradictory effects. While Atkinson et al. (2009) find that the performance-related pay scheme implemented in England did improve test scores and value added increased on average by about 40% of a grade per pupil, Ballou (2001) shows that in the case of the U.S. efforts to implement merit pay in public education have generally been unsuccessful, mainly because of the opposition from teachers and teachers' unions. Lavy (2015) reports persistent gains in labour market achievements of students whose teachers were exposed to pay-to-performance schemes in Israel. Since 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 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, the presence/absence of monetary incentives in the teacher profession induces self-selection of different individuals.<sup>3</sup> Whether these two policy instruments are substitutes or complements is hard to judge, since self-selection occurs on unobservable characteristics, which in turn can be correlated to (unobservable) teacher's quality. Merit-pay wage policies should attract people expecting to benefit most from such a scheme, but whether they are abler or/and greedier than average is difficult to gauge: as a consequence, it is almost impossible to predict which can be the overall effect on students' achievement, since the "selection" and the "incentive" effects may work in opposite directions. If therefore it is impossible to derive uncontroversial predictions about which 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 teachers recruiting and rewarding policies to identify which of them are most effective in raising student achievements.

### 3. Data and Descriptive Statistics

The growing interest for educational outcomes in the last decades increased the demand for high quality data by both academics and policy makers. International comparable surveys monitoring pupils' achievements at different grades have become more and more

---

<sup>3</sup> "...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." (ibidem, p.F256).

widespread in developed and developing countries. Namely, reading and literacy proficiency are tested at 4<sup>th</sup> grade by PIRLS, numeracy proficiency is tested in two subsequent stages of the learning path at 4<sup>th</sup> and 7<sup>th</sup> grade by TIMMS, while all the three competences (including scientific knowledge) are assessed at 10<sup>th</sup> grade through PISA. Although these surveys are informative in a cross-country perspective (Volante, 2017), they do not allow a complete description of the learning environment for students. Information on teaching practices are self-reported by either students or teachers/school principals while only essentials teachers' characteristics are recorded, making it difficult to study the direct contribution of teachers to learning. Therefore, for our analysis we combine survey data on pupils, teachers and schools belonging to the first cycle of the educational system with a newly created dataset of country-level time-varying institutional features relevant for primary teachers over the last century.

Microdata on pupils' achievements are drawn from the three waves of the PIRLS assessment run every 5 years from 2001 to 2011, as available in early 2017, covering about 70 country/state/region entities with legal autonomy in educational policy making. Information about the country-level institutional setting for teachers have been collected from secondary data sources and assembled in a final dataset that covers 56 countries/regions surveyed by PIRLS with a yearly frequency over the period 1947-2011. Let us shortly describe the content of these datasets and to present basic descriptive statistics.

### **3.1 Microdata on pupils', teachers' and schools' characteristics**

The PIRLS assessment is conducted by the International Association for the Evaluation of Educational Achievement (IEA) and it tests reading literacy of 4<sup>th</sup> grade pupils. The 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 for reading, the reading behaviours and attitudes. Student performance is measured by test scores in reading literacy, standardized 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 34 countries/state/regions in 2001, 43 in 2006 and 58 in 2011; however in Table 1 we report the list of countries included in our analysis, restricted to countries/waves with non missing values for any of the variables used in the empirical models (18 in 2001, 40 in 2006 and 45 in 2011).<sup>4</sup>

Test scores are nationally representative. National samples were drawn through a two-stage stratified sampling design. First, 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 pupils participated to the assessment. Together with students' reading achievement, the survey collected background information on students, parents, teachers,

---

<sup>4</sup> We also replicate the analysis reported in the main text over the largest available sample, without finding any significant differences. Results are available upon request.

schools and curricular activities. Questionnaires were administrated to the tested students, to their parents, to their reading teachers, and to their school principals. Teachers' information refers to the main or unique reading teacher of the class, though we ignore whether the same teacher taught the same pupils during the previous grades.<sup>5</sup>

Table 2 provides summary statistics of core variables used to perform the empirical analysis. Beside the reading test scores, we group individual socio-demographic features, school and teacher characteristics. We restrict the analysis to those pupils with a complete record of data for their parental background as well as for their schools and teachers. Among individual observable features potentially responsible for 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. Schools features refer to the size, the geographical location or the presence of IT resources as well as a library for pupils. Among teacher characteristics we focus on gender, age, educational level and tenure.

### 3.2 Institutional setting relevant for teachers

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

Namely, we select the following dimensions as those most likely to affect teachers' efficacy and students' performance: (i) having a compulsory training period during or immediately after the Teacher Educational Programs as a requirement for teaching; (ii) passing a standardized 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 early stages of the teacher career; (v) having a mentoring program for teachers; (vi) having an official process to license or certify teachers by one of the following institutions: the ministry, a National/State Licensing Board, the Union, Universities/Colleges; (vii) receiving a specific preparation on teaching techniques. For each dimension but the fourth one, we construct a dichotomous indicator taking value 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 in charge to different institutions as well as the evaluation process through certifications. Therefore, we include also a variable accounting for the total number of certifications required to teach at 4<sup>th</sup> grade. Finally, we construct a synthetic *index of selectivity* as the mean

---

<sup>5</sup> The structure of the dataset is nested, with 4 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 may argue that teacher quality declines with the number of classes taught (this happens only in 758 classes over 30121).



of the previous dichotomous indicators for every country-year. A higher index means a more stringent selectivity for primary school teachers which should translate in higher quality of the teaching staff and in higher pupils' performance.

Since also prevailing wages could affect the quality of the pool of teachers, we complement data on the reference institutional setting for primary schools with information about 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 to other professions requiring similar qualifications in terms of education. Higher relative pay for teachers should attract better quality candidates and/or enhance their quality in terms of skills and motivation. Hence, from various issues of OECD's *Education at a Glance* we collect the ratio of primary teacher salary to GDP per capita for each wave of PIRLS. Unfortunately, information on wages is available only for a subsample of 31 OECD countries. This reduced sample will be the smallest one on which we will perform our empirical analysis.

Every country-level indicator is then associated to pupils by country/waves. Summary 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, sufficient variation emerges within countries. Most indicators exhibit positive correlation among them, suggesting that these dimensions complement and reinforce each other's among educational systems. The few exceptions of negative correlation are not statistically significant.

### **3.3 Reforming process of the teacher profession**

We also collect detailed information about the reforming process that took place in each country over previous decades with respect to the teacher profession, focusing on the selection and the rewarding systems, as factors that affect the quality of teachers. Namely, we were able to identify four reforming areas: (i) reforms affecting teachers' recruitment process; (ii) reforms on the pay scheme; (iii) reforms on the working conditions; and (iv) reforms affecting the 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 include those changing the pre-requisite criteria, through changes in the minimum marks to enter teacher colleges, in the level of educational attainment or in the prerequisites for the teacher certification or licensing. Reforms of the working conditions refer to changes in the working hours, in legal rights for special leaves or in the 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, often under the pressure of teachers' unions. Finally, reforms of the retirement rules include legislative changes in the retirement entitlements (specific for teachers, since we are interested to the incentive mechanisms of teachers' selection vs. other careers) such as the pre-requisites 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 ILO’s International Labour Standards Department<sup>6</sup>. 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 use also country-specific descriptions of national education systems as well as thematic studies on specific institutional features. Finally, we double-checked our data by direct contacts with national experts on the field.<sup>7</sup>

Exploiting these data sources, we are able to identify the exact year of implementation of each reform, as well as the direction of the change that took place (i.e. whether it is favourable or not for teachers). As a result, we assemble an original dataset containing yearly observations over the 1947-2011 period for all countries surveyed in PIRLS on the implementation of legislative changes in the previous four dimensions of the teacher profession. Whenever, in a given year, we record a change in a specific dimension we assign a value one, while if no changes occur we assign zero. When legislators have repeatedly reformed a specific dimension over the sample period, we created step dummies, which are then summed up over the years, with a final normalization yielding a unitary range of variation. For all the dimensions we construct the indicators based only on the direction of the legislative change so that an increase/decrease in the variable refers to a legislative change that is favourable/unfavourable to teachers. We end 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 a more generous wage allowances for primary school teachers. Finally, the fourth one captures the stringency of retirement conditions and includes the retiring allowance, the severance pay and the 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 teachers surveyed by PIRLS according to the year when they entered the labour market. Consider the case of “Teacher recruitment reform

---

<sup>6</sup> The database lists and classifies all the legislative actions about several fields broadly related to the labour market and working conditions, ranging, among the others, from Employment security to Maternity protection or Elimination of forced labour. We focused on the categories “Education, vocational guidance and training”, “Conditions of employment”, “Conditions of work”, “Specific categories of workers => Teachers”. For each action, several information is 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 select those relevant to our scope and classify them according to the four broad categories described in the text.

<sup>7</sup> 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, 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, biasing downward our estimates.

indicator” in Italy as an example. Sources of data report more stringent reforms over this dimension in 1987, in 1988 and in 1990. We therefore construct a variable, which is zero before 1987, 1/3 in 1987, 2/3 in 1988 and 1989 and 1 afterward. Every teacher entering the profession before 1987 gets a 0 value for this reform, those entering in 1987 get 1/3 and so on, those in 1988-1989 receive 2/3, while a value of 1 is associated to teachers hired more recently.

Table 5 reports summary statistics of our original reform variables not yet matched with pupils and teachers PIRLS micro data, while in Table 6 we report the pair-wise correlation matrix among them. Notice that the decomposition of the standard deviation into the “between” and the “within” components indicate that there is sufficient variation within countries.

#### 4. Empirical strategy

The aim of the empirical strategy is to identify whether some policies intended to *attract*, to *select* and/or to *motivate* good teachers can improve pupils’ performances in primary schools. In order to test whether teachers matter for pupils’ performance one should correlate pupils’ achievement with measures of teacher quality. However, we have already reviewed the problem of measuring teachers’ quality: observable characteristics of teachers are weakly correlated with student achievement and the reverse strategy of inferring teachers’ quality from observed student achievements is only valid when either students are randomly allocated to teachers (inapplicable for countries where there is explicit or implicit streaming) or one possess longitudinal samples where repeated observations of different student cohorts are exposed to the same teacher (Rivkin et al., 2005). This strategy reveals even more complicated when we consider that pupils 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 *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 in order to identify policies that may reveal effective either because they attract/select better teachers or because they solicit higher level of effort. In addition, we focus only on the effect of the main instructor in charge of teaching reading at 4<sup>th</sup> grade to pupils. Unfortunately, the survey does not collect information about teachers to whom pupils have been exposed in previous grades, if different from the current one, making it impossible to distinguish between the two different effects.

The empirical analysis proceeds in three subsequent steps. We first review the standard correlations of pupils’ achievement with individual, class and school characteristics, as well as characteristics of the teachers responsible for reading. For each pupil  $i$  associated to class/teacher  $j$  in school  $s$  of country  $c$  surveyed in year  $t$  we estimate a standard educational production function for her 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  $X_{it}$  associated to the pupils 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 to the class can be decomposed into two sub-vectors: the first one,  $X_{ijt}$  includes contextual class effects computed excluding the considered pupil (like share of female in the class, average age in months, share of immigrants and/or of pupils speaking a different language at home, an index for household educational resources, average educational attainments among the parents in the class); the second sub-vector  $T_{j\tau t}$  contains information regarding the main or unique reading teacher of class  $j$ , who entered the labour market in year  $\tau$  and surveyed in year  $t$ : gender, age (in 10-years intervals), tenure (years), educational attainment (being graduate and having an official certification for teaching). The third vector  $Z_{st}$  includes school characteristics like location (urban/rural), average teachers' tenure in the school, availability of library and IT technologies, share of disadvantaged students in the school and school size.  $\delta_t$  and  $\delta_c$  are wave and country fixed effects, while the idiosyncratic error component, clustered at class level, is  $\varepsilon_{ijsct}$ . School fixed effects can also be considered instead of including school characteristics, and in such a case country fixed effects are removed.

The second step consists of the analysis of the contextual effects affecting primary school teachers at country level, and possibly changing from one survey to the other. Hence, we introduce institutional features and policies targeted to teachers and teaching activities and 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  $W_{ct}$  containing information about the institutional design of the teacher recruitment process at country-year level as well as their relative wage. Obviously, this specification has two main limitations. First, since institutional features are observed only over one decade, they show limited (if any) variation within country over time. Hence, the estimated effects might be confounded with other institutional elements at country level, which are not explicitly accounted for, and prevent the inclusion of country fixed effects in the model. Second, since 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 in the profession. We can only estimate the effect that contemporaneous correlation between relevant settings for teacher and pupils' performance. As such, they are likely to affect directly 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,

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

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  $R_{ctt}$  of the implemented reforms in country  $c$  at time  $t$ . The match of reforms to teachers according to their year of entry in the labour market allows identifying the effect of policies by comparing students' achievement in classes taught by "treated" teachers against classes taught by "non-treated" teachers acting as control cases. In fact, by matching reforms to teachers based on their age and experience we can distinguish those who were affected from those who were not. As

an 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 exiting teaching schools in the same year were forced to undertake three additional years in order to obtain the degree. Thus, other things constant, we can test whether pupils in classes with teachers hired before 1990 exhibit a worse performance compared to pupils taught by teachers hired after 1990 (presumably with a BA degree – for 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 since age/tenure effects are separated by observing individuals in the same labour market with the same age/tenure but matched with different set of teacher policies since they are observed in different time periods (thanks to the availability of three surveys spanning a decade). Finally, the repeated cross-section nature of our data allows distinguishing the age effect and the cohort one.

Our identifying assumption is that reforms in teacher policies (especially recruitment ones) affect beginner teachers only, leaving already tenured teacher unaffected (i.e. any imitative behaviour can be considered as negligible). Similarly, we consider that reforms regarding pay and retirement rules affect all teachers, but at a different degree of intensity, which is stronger the younger is the teacher (i.e. the smaller is the time period between entry in the profession and reform, via an effect on 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 the comparison of the magnitude of their impact across countries and over time. As such, our variables capture the frequency/intensity of the reforming activity of subsequent governments, vis a vis the teachers within each country.<sup>8</sup>

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. In order to have a consistent sample across all specifications, we rely on 56 countries for which we have complete information in all variables included the main model except wage that is available only for the subset of OECD countries. We run our analysis in the full and in the restricted sample to test whether results are sensitive to the restriction to rich/OECD countries sample.

## 5. Results

In this section we present the main results proceeding by gradual steps. Estimates for the educational production function described in equation (1) are presented in Table 7, where we correlate pupils' performance with teachers' characteristics, controlling for pupils' demographics, socio-economic background, 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 sub-sample of countries for which data on teachers' salary are available, that is, OECD countries. In all the specifications standard errors are clustered at school level.

---

<sup>8</sup> Similar strategy has been pursued by Braga et al. (2013) while studying the impact of educational policies onto educational inequalities.

Regardless of the specification, in line with previous evidence, female students outperform their male counterpart by an average of 12 points. Point estimates show a very small negative effect of age that probably captures lower skills of students repeating the year. The socio-economic background of students is positively correlated with their attainment: pupils 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 in disadvantage. It is important to note that the PIRLS survey directly tests linguistic competences 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 strengthen the effect of the corresponding individual feature, suggesting 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 pupils' performance. The signs of gender, age and tenure are in line with previous researches (see 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 of the 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 age, tenure (in terms of years of continuous activity in primary school) does not always influence pupils' performance.<sup>9</sup> Instead, differently from other studies, we find a statistically significant effect of teacher education (Chingos and Peterson (2011)). In particular, having a teacher with at least tertiary education increases on average pupils' performance in the standardized tests by 4 to 7 points depending on the specification. Increasing teachers' educational attainment seems to be a driver to enhance primary pupils' performance. The same effect is not found in previous studies mainly focused on higher levels of the educational system. However, this result is not surprising since, in the past teachers entered primary schools after completing teaching schools, corresponding to secondary/vocational education. The completion of tertiary education was typically required for higher school levels. Over time, requirements to enter the profession have increased and hence, on average, a differential effect emerges between teachers with 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 often the teacher profession is less attractive for more educated and qualified new entrants to the labour market.<sup>10</sup>

Having identified the basic determinants of pupils' performance, we go ahead in our analysis by exploring the role of country specific institutional features in shaping average test scores, as described by equation (2). Results are presented in Table 8. In all specifications, we control for the same variables at individual, teacher class and school level included in previous estimates. School fixed effects cannot be considered in the model

---

<sup>9</sup> This is important in our perspective, since we match teachers and reforms basing on this variable. Its limited significance in this regression reduces the risk of spurious correlation with the reform variables.

<sup>10</sup> To see whether country specific heterogeneity drives these results we estimate equation (1) for each country. The Netherlands and Quebec are the two countries that account for the negative effect of tertiary education on pupil's performance. Instead, the age effect is heterogeneous across countries.

since institutional features are country invariant in any survey year. In Column (1) we report 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, obtained as a normalized mean of the previous eight indicators. Column (3) replicates the specification of Column (1), with the inclusion of a pay index. Columns (4) and (5) include the same models as in Columns (1) and (2), restricted to the sub-sample of OECD countries, while the remaining columns show the effects of pay alone (6) and the interaction between pay and selectivity (7).

All institutional characteristics included in the model are positively correlated to students' performance. Countries characterised by more selective recruitment or better-structured training period enhance pupils' competences. 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 profession. Similar role is exerted by the completion of a probationary period or requiring a license to certify primary teachers. More specifically: in countries where teachers receive a specific preparation on how to teach reading, pupils obtain 29 additional points than the average in the full sample (18 in the restricted one); similarly, in countries envisaging a probation period for primary school teachers the average reading test scores are from 10 to 17 points higher than elsewhere. At last, having to pass an examination before starting the career is also positively associated to pupils' achievement, with an average effect ranging between 5 and 7 points. The certification process is the only institutional feature sensible to the estimation sample: having an official certification process to license primary school teacher and increasing the number of required certifications is positively associated to pupil performance in the full sample (including middle-low income countries), while it reverses its sign when restricting to high-income countries. This could be taken as indicative that simple accreditation, without adequate training, does not produce increase in quality of teaching, especially where supply of potential teachers is relatively abundant.

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 specialization for primary teachers in a country. Such specification allows testing whether with our previous estimation we were capturing some other country specific effect and not only the institutional setting of reference for primary teachers. Results presented in Columns (2) and (5) corroborate, in both samples, those obtained when each single dimension is taken into account. Overall, on average, pupils perform better in countries where teachers are more selected and specialized than elsewhere.

In Column (6) we analyse the effects exerted by remuneration including as 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 to others. A higher ratio should enhance teacher quality both in terms of skills and motivation. Indeed, the coefficient associated to the pay index is positive and significant, showing a strong positive correlation between the reward of teachers and the performance of their pupils (in line with the findings of Dolton and Marcenaro-Gutierrez 2011). The relative remuneration appears as one of the institutional features enhancing pupil performance via attracting better

candidates among entrants into the teacher profession but also as a tool to retain and motivate them during their career.

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. Teachers selection and reward appear as substitutes, that is, the effect of wage on pupils' performance is attenuated by an increase in selectivity and the effect of selectivity on performance declines as wage increases. From a policy perspective, since 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 spurious correlation, though some of these dimensions change over the different survey years. Mainly for this reason we have resorted to our preferred strategy, represented by the model described by equation (3). In Tables 9 and 10 we test this hypothesis that considers the effect on pupils' performance of a change in the institutional setting relevant for primary school teachers just before starting the career. All models are estimated including students', 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 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 sub sample of countries where information on remuneration were presents in order to control for the effect exerted by the pay level. The former are presented in Columns (1) and (4), the latter are presented in the remaining columns.

The two reforming areas potentially enhancing pupil performance through teachers' quality are the introduction of a more selective recruitment process and more favourable reward policies, a sort of "warm glove" attitude toward teacher, including more generous pay conditions, more favourable working conditions and/or more advantageous retirement policies. A more selective or targeted recruitment process is designed and implemented to select well-qualified candidates with 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 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 policymakers.

The selectivity of the recruitment policy is robust to the sample as well as to the inclusion of fixed effects to account for unobservable heterogeneity within countries over time. Also when considered together with other reforms it still affects pupils' performance (Column 1) even controlling for the wage level (Column 2). Instead, the effect of reforms of the remuneration 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 the improvement in working conditions decreases on average the performance of their students signalling that such policies tend to attract less motivated and/or individuals with conciliation problems



due to housewifing duties (for example female teachers that find this profession easier to combine with caregiving in their households). Negative effect is found when considering changes in the retirement schemes (for the OECD restricted sample) indicating again that the possibility of early retirement may lower the level of individual motivation, possibly due to short-sightedness. Results hold irrespectively from 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 to raise students' 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 tertiary or post tertiary educational level. Estimates are presented in Table A.1.1 in Appendix, where in each line we report point estimates for 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 non-graduate teachers while the effect for graduated 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 non-graduate teachers in Column (2) and their difference in Column (3). In Columns (4) to (6) we replicate the same structure on the restricted sample for which information on pay is available. Though effects are not the same along all 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 to better students' performance in countries where the institutional setting is more selective. In the subsample of the OECD countries, when controlling for the pay level, the opposite situation manifests. In rich countries, the institutional framework seems to be more effective for less qualified teachers, suggesting that part of the selection and/or the attractiveness is exerted by attending university courses. Results are robust to the sample as well as to the inclusion of the pay index.

We explore the heterogeneity also by interacting each reform with the level of education of the teacher. Estimates are presented in Table A.1.2, showing a clear trend: the effects of all reforms are stronger for non-graduate teachers. More in detail, stricter recruitment processes and retirement reforms have a positive effect for non-graduate teachers and a negative (or insignificant) effect for graduate. Pay reforms are effective to improve only performances of (children taught by) non-graduate teachers. Working conditions reforms affect non-graduate teachers less negatively than graduate ones.

In all previous specifications, we focused on the intensity of the reforming process occurred before each teacher entered the labour market. Such reforms should affect the quality of applicants and hence the subsequent performance of their pupils. However, also having been exposed to reforms throughout the career could influence the incentives to be effective in teaching. Therefore, in Table A.1.3 we study whether the intensity of the

reforming process during the career has an effect on teacher quality and translates in different student' achievements. In particular, for each one of the four reforming areas, we identify the number of legislative changes hitting a given teacher after his entry in the school system up to the date of the survey when pupils' competences are tested. All models are estimated including student, teacher, class and school characteristics together with wave and country fixed effects. Teachers exposed to more changes in recruiting process or in pay conditions are more effective in improving students' performance. The pay structure and working conditions appear as two effective tools to influence incentives for primary school teachers (though in an opposite direction). On the contrary, a higher number of reforms in the retirement process has no clear effects on pupil performances, suggesting that such changes do possibly change incentives throughout the career.

An implicit and untestable assumption of our identification is that the selection into the profession is more affected by 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 from individual level information. Hence, we test whether the effects of the institutional setting dissipate over time by running separate analysis according to the teacher tenure. As shown in Table A.1.4, based on the synthetic index for the selection process, enhancing selectivity is more effective with 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 less tenured teachers. Finally, Table A.1.6 shows that also the reform activity has a heterogeneous effect according to tenure. In particular, while reforming working conditions and retirement rules have a negative effect that is mostly homogenous over tenures, the effects of recruitment reforms and salary reforms raise their intensity for longer tenured teachers than for the others.

In addition, to better characterize the role played by the pay level in attracting better workers as primary school teachers, we recover information about the (1) ratio of salary at top of scale to starting salary, (2) years elapsing from starting to top salary and (3) salary per hour of net contact (teaching) time after 15 years of experience. Although not fully comparable since data are available only for a subset of countries, estimates are in line with those discussed in the main text. In details, sign and significance of the top/least salary ratio is the same as our main indicator, that is positive and always significant. The steepness of the salary curve is positively correlated to the performance of pupils, but the significance disappears once we control for the top/bottom wage ratio, indicating that what matters is the overall salary profile, more than the speed to reach the top. Finally, hourly wage is negatively correlated to pupils' outcome, but also 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 pupils' outcome increases with total income rather than with hourly wage. Stated differently, keeping fixed the hourly wage, the effectiveness of teachers increases is positively correlated to hours worked.

Finally, though not reported in the text, we also collect data on the statutory teacher wage at starting career, the average wage after 10 and 15 years of experience and the wage at top of the pay scale. In order to match wages and teachers at different seniority levels, we interpolate linearly the four points on the wage scale, assuming then that wages smoothly increase over time. The likely introduction of 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 results indicate that a one percent increase in wage translates into a 3 points higher test score, that is .6% of the standardized 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 non-random allocation of institutions and reforms over time and across countries. We therefore perform a falsification test in which we estimate our basic equations (Equation 2 and 3) using as regressors a vector of randomly generated institutional features prevailing in the three waves or when teachers entered 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. Point estimates for both the single institutional features and for the synthetic indices are no longer statistically significant and when significant they obtain the opposite sign (with the only exception of the Examination dummy). Overall these falsification tests suggest 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 reforms to teachers in two different ways. First, we randomly assign reforms across countries within tenure cohorts; second, we assign 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, in order to check whether the results are driven by the behaviour of a single country, we re-estimate our baseline regressions excluding from the sample one country at a time or for each country separately. Although not reported in the paper for space reasons, both for institutional features and reforming activity, results continue holding and are not driven by the behaviour of a single country.<sup>11</sup> Similarly, we perform separate regressions for different geographical areas to test whether specific groups of countries drive our results. In particular, we perform separate regressions for Europe, formerly planned economies, North America, Latin America, East Asia, Middle East and North Africa, Oceania and Sub-Saharan Africa. No specific patterns emerge using regional disaggregation.

## 7. Discussion and Conclusions

This paper provides new evidence on the effect of teacher quality on pupils' performance in primary school. Based on international standardized tests for literacy conducted on 4<sup>th</sup> grade pupils, and using variations in the institutional setting, our analysis shows that teacher

---

<sup>11</sup> Results are available upon request.

quality matters. The two main channels to enhance teacher quality are selectivity at the entry of the profession and reward. Reward can be disaggregated into compensation, working conditions and early retirement eligibility, but they emerge as statistically significant, though with an opposite sign: offering higher pay or lighter working arrangement (possibly accompanied by advantageous retirement conditions) seem to work in opposite directions when 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 education budgets.

Our results suggest that policies could effectively enhance school quality. Indeed, some of the dimensions we consider are actually introduced by policymakers to better train teachers providing them with adequate skills. Better formal professional training shows up as a key factor to be successful and effective in class teaching practices, to improve learning and to assure high-quality standards. Specifically, on the one hand, the presence of examinations or licences is a screening device to assess competences and capabilities of aspiring teachers. On the other hand, the purpose of a probationary period is to ascertain whether the conduct and work performance of the teachers meet the standard expected by their employers before the full rights and responsibilities of on-going employment are confirmed. The probationary period also provides support and feedback to teachers at the beginning of their career. 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 required to perform well in teaching. Effective pre-service 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 four main areas of human resource management: recruitment, working conditions, pay and retirement conditions. Our identifying assumption consists of reforms mostly hitting teachers at the time of their entry in the profession, and gradually dissipating afterwards. As such reforms are salient in attracting better-qualified and more motivated candidates. We find that selection and generous pay are both effective tools to attract good aspiring teachers, but when jointly considered they appear as substitutes, since 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 aim to improving pupils' performance. In addition, when targeted to primary school teachers, these reforms could also be effective in enhancing the overall quality of 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, also including health insurance, pensions or fringe benefits. In these cases, wages differentials over time and across countries capture only one dimension of the rewarding scheme that could influence teachers' productivity. Our analysis also neglects other dimensions of the educational process, ranging from school infrastructure to extra-curricular activities or school timetable. These dimensions,

whether not proxied 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* (2<sup>nd</sup> 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.
- Schäcter, 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			Total	OECD salary
		2001	2006	2011		
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
<i>Individual students' characteristics</i>					
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
<i>Schools' characteristics (weighted by students)</i>					
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 4 <sup>th</sup> grade students	278099	0.36	0.61	0	48
<i>Average class students' characteristics (weighted by students)</i>					
Share of females	278099	0.5	0.18	0	1
Class size	278099	25.75	7.69	1	119 <sup>12</sup>
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
<i>Teachers' characteristics (weighted by students)</i>					
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

<sup>12</sup> Even if this value can seem unrealistic, the 99<sup>th</sup> 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)
<i>Pupils' characteristics</i>						
Female	14.17*** [0.280]	13.04*** [0.278]	13.46*** [0.248]	11.43*** [0.358]	10.92*** [0.332]	10.76*** [0.323]
Age in months	-0.31*** [0.027]	-0.18*** [0.027]	-0.13*** [0.025]	-0.43*** [0.037]	-0.29*** [0.034]	-0.36*** [0.036]
Different language spoken at home	-7.38*** [0.338]	-7.25*** [0.341]	-7.26*** [0.322]	-10.99*** [0.459]	-11.52*** [0.428]	-10.76*** [0.429]
Index of home educational resources: Medium	-20.39*** [0.437]	-27.57*** [0.453]	-21.16*** [0.421]	-22.56*** [0.555]	-24.15*** [0.538]	-23.49*** [0.520]
Index of home educational resources: Low	-32.12*** [0.882]	-43.87*** [0.893]	-32.65*** [0.845]	-42.62*** [1.415]	-49.32*** [1.361]	-44.92*** [1.353]
Highest Parental Education: Tertiary or post	48.14*** [0.782]	42.00*** [0.832]	47.41*** [0.742]	48.75*** [1.200]	46.49*** [1.176]	47.08*** [1.151]
Highest Parental Education: Post-secondary	35.71*** [0.767]	34.75*** [0.802]	35.82*** [0.723]	38.24*** [1.209]	37.14*** [1.170]	37.19*** [1.155]
Highest Parental Education: Upper secondary	24.02*** [0.732]	23.08*** [0.778]	23.94*** [0.689]	27.32*** [1.169]	26.42*** [1.142]	25.90*** [1.122]
Highest Parental Education: Lower secondary	8.72*** [0.725]	11.56*** [0.751]	9.25*** [0.682]	8.73*** [1.189]	10.77*** [1.152]	8.50*** [1.134]
<i>Teachers' characteristics</i>						
Female	4.13*** [1.057]	5.90*** [1.274]	4.73*** [0.918]	3.48*** [1.036]	2.84*** [1.002]	2.42*** [0.859]
Age group: Under 25	6.37* [3.414]	20.68*** [3.729]	5.58** [2.839]	-0.04 [3.537]	10.61*** [3.515]	-1.65 [3.023]
Age group: 25-29	5.17* [2.667]	15.63*** [3.079]	8.63*** [2.214]	-3.74 [2.981]	8.57*** [2.710]	-0.95 [2.322]
Age group: 30-39	4.98** [2.254]	9.66*** [2.552]	6.75*** [1.859]	-0.57 [2.631]	6.43*** [2.426]	-0.22 [2.056]
Age group: 40-49	4.80** [1.915]	11.65*** [2.171]	4.76*** [1.585]	-0.11 [2.347]	4.73** [2.194]	-1.00 [1.820]
Age group: 50-59	4.56*** [1.716]	11.87*** [1.975]	3.55** [1.423]	1.41 [2.134]	5.29** [2.062]	-0.39 [1.642]
Tertiary education	7.11*** [2.044]	5.07*** [1.622]	4.04*** [1.567]	5.80** [2.835]	-11.50*** [1.843]	1.19 [1.869]
Tenure	0.10 [0.060]	0.18** [0.090]	0.20*** [0.067]	-0.05 [0.063]	0.13 [0.081]	0.03 [0.068]
<i>Class' characteristics</i>						
Female (%)	22.91*** [2.924]	7.54*** [2.459]	12.47*** [1.586]	14.66*** [3.422]	7.29*** [2.408]	5.79*** [2.123]
Age in months (average)	-0.19 [0.230]	0.94*** [0.087]	2.33*** [0.131]	-0.39 [0.258]	1.66*** [0.085]	0.41** [0.189]
Different language spoken at home (%)	-10.75*** [2.572]	-7.69*** [2.056]	-11.10*** [1.681]	-18.83*** [3.146]	-29.29*** [1.873]	-19.06*** [1.958]
Index of home educational resources: High (%)	46.22*** [7.166]	206.52*** [7.041]	47.73*** [5.723]	41.99*** [10.186]	122.10*** [8.122]	66.28*** [7.722]
Index of home educational resources: Medium (%)	27.09*** [6.284]	81.23*** [5.977]	18.33*** [4.790]	31.30*** [9.517]	91.29*** [7.461]	44.13*** [7.191]
Highest Parental Education: Tertiary or post (%)	96.46*** [6.756]	39.80*** [6.435]	95.53*** [4.896]	47.57*** [7.593]	34.06*** [6.445]	33.58*** [5.731]
Highest Parental Education: Post-secondary (%)	78.03*** [6.706]	97.56*** [6.167]	89.21*** [4.914]	35.46*** [7.670]	36.47*** [6.307]	29.00*** [5.733]
Highest Parental Education: Upper secondary (%)	53.48*** [6.475]	70.26*** [6.076]	61.52*** [4.648]	19.15*** [7.430]	21.80*** [6.193]	6.89 [5.643]
Highest Parental Education: Lower secondary (%)	44.24*** [6.584]	107.18*** [6.010]	58.72*** [4.804]	13.33* [7.824]	52.85*** [6.563]	15.99*** [5.998]
Number of students	0.93*** [0.123]	-0.64*** [0.086]	-0.29*** [0.066]	0.62*** [0.147]	-0.11 [0.085]	0.03 [0.064]
<i>Schools' characteristics</i>						
Total number of students		-0.00** [0.002]	0.00** [0.001]		-0.01*** [0.002]	0.00 [0.002]
Students in the 4th grade		0.12*** [0.012]	0.02** [0.008]		0.13*** [0.017]	0.00 [0.016]
Urban		4.69*** [0.872]	1.62*** [0.596]		2.57*** [0.689]	0.22 [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 4 <sup>th</sup> 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 Induction Program (dummy)	0.85		4.54***	6.11***			
	[0.885]		[0.827]	[0.837]			
Specific preparation in how to teach reading (dummy)	29.33***		16.85***	19.48***			
	[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 4<sup>th</sup> grade, a dummy for schools in urban areas, share of disadvantaged pupils (in 4<sup>th</sup> grade), average tenure within the school, a dummy for the presence of a library, number of computers for pupils in the 4<sup>th</sup> 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*** [1.174]	2.77*** [1.069]	2.62** [1.066]	3.19*** [1.211]	2.21** [1.055]	2.08** [1.053]
Working conditions reforms	-2.89** [1.233]	-1.97 [1.211]	-1.85 [1.211]	-2.53** [1.145]	-1.88 [1.184]	-1.82 [1.184]
Salary reforms	-1.01 [1.084]	0.35 [1.043]	0.21 [1.047]	-1.50 [1.018]	-0.15 [1.007]	-0.24 [1.011]
Retirement reforms	-1.46 [1.338]	-2.53** [1.209]	-2.43** [1.210]	-1.52 [1.329]	-2.52** [1.204]	-2.45** [1.204]
Pay index		13.10* [6.859]			included	
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.

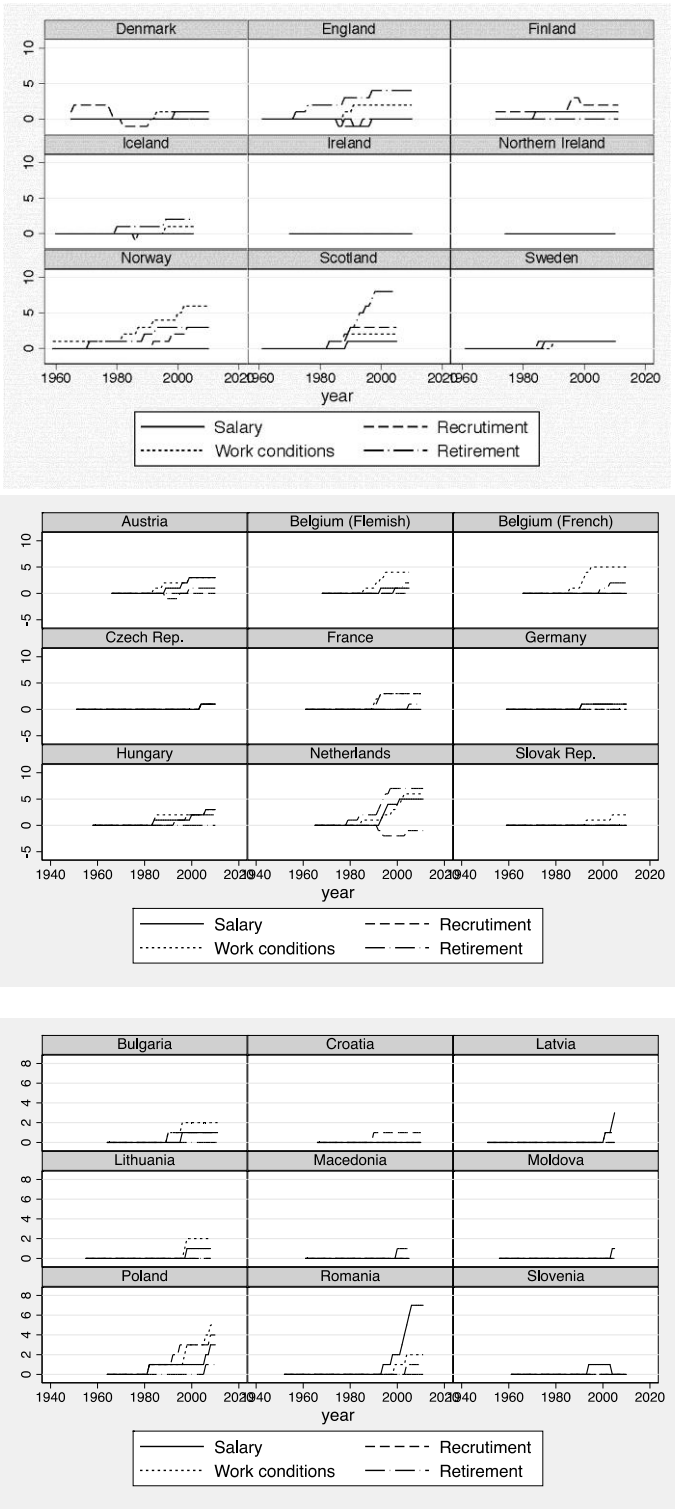
**Table 10. Reforms on teachers' career and pupils' test scores.**

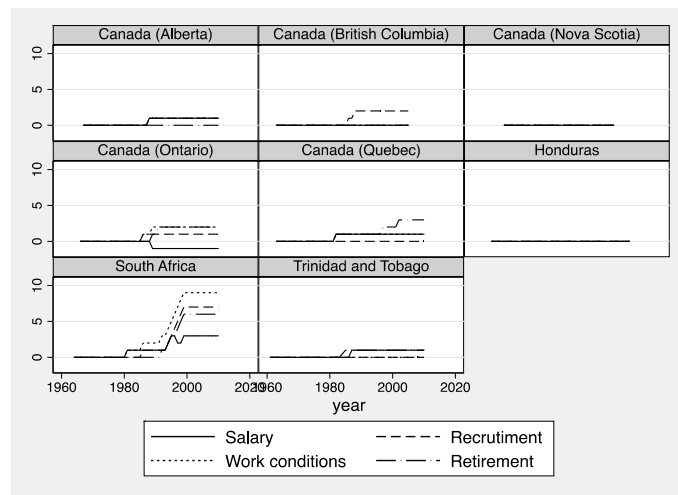
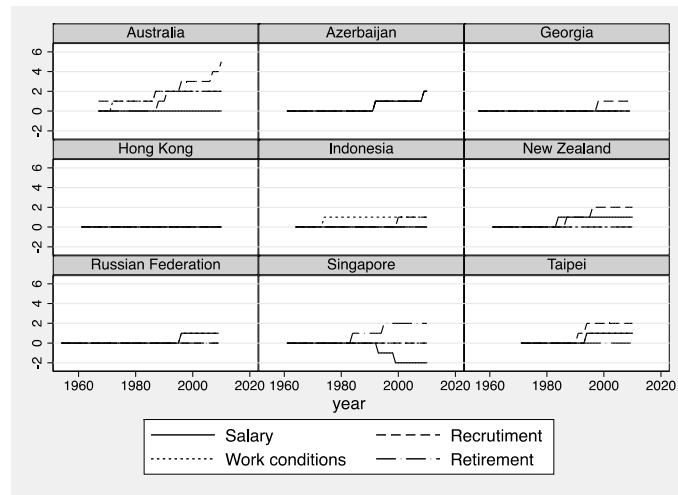
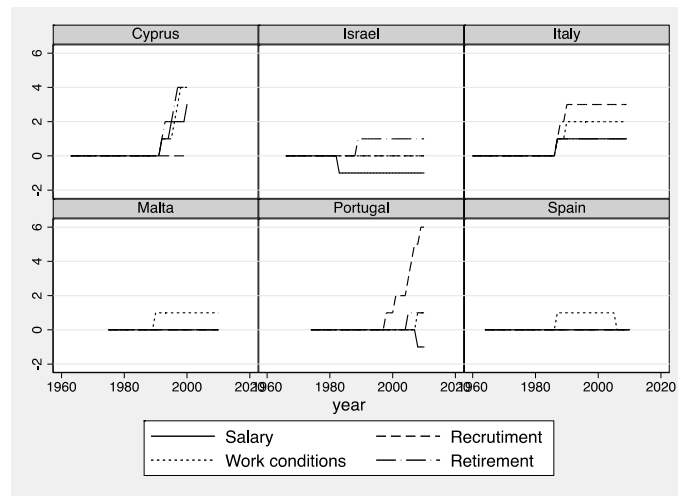
Sample	Main	Reduced OECD		Main	Reduced OECD	
Variables included		All			One by one	
Dependent variable:						
Standardized Test score	(1)	(2)	(3)	(4)	(5)	(6)
Recruitment reforms	4.17*** [1.035]	4.35*** [0.825]	4.43*** [0.830]	-0.34 [1.098]	3.38*** [0.808]	3.50*** [0.819]
Working conditions reforms	-13.85***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

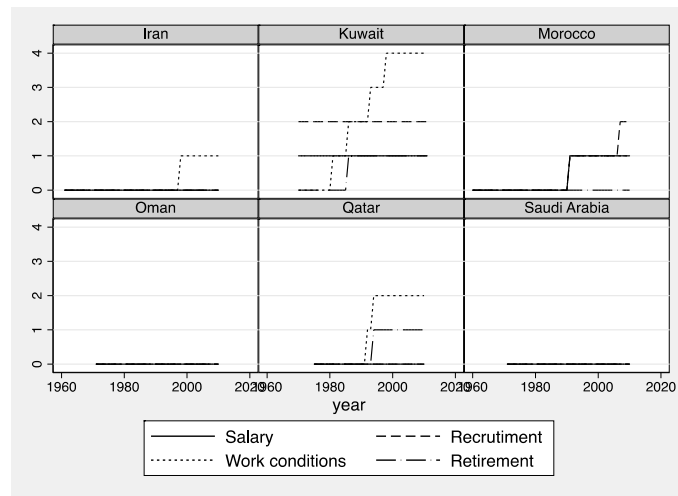
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1, robust standard errors clustered at school level in brackets. All regressions include pupils, teacher, class and school characteristics as listed in Table 7. Wave fixed effects are included. Coefficients for pay index are always positive and significant at 1% level.

FIGURES

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









## 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)
<i>Teachers' education:</i>	Graduate	Non graduate	Difference	Graduate	Non graduate	Difference
Any training before teaching (dummy)	-6.68**	-17.14***	10.46***	-15.93***	-10.27***	-5.66**
	2.889	3.320	1.850	2.284	3.268	2.513
Passing an Examination (dummy)	12.38***	6.99*	5.39***	49.35***	66.02***	-16.67***
	3.681	3.961	1.760	9.067	9.274	1.544
Length of probationary period	4.32	0.27**	4.05	-5.95**	1.04***	-6.99***
	2.770	0.105	2.686	2.735	0.107	2.654
Probationary period (dummy)	9.20**	-6.47	15.67***	30.44***	48.55***	-18.10***
	4.492	4.681	1.698	4.841	5.095	1.617
Completion of a Mentoring or Induction Program (dummy)	-4.54*	-21.74***	17.20***	-4.39*	9.76***	-14.15***
	2.328	2.903	2.206	2.351	3.173	2.507
Specific preparation in how to teach reading (dummy)	1.10	-7.28*	8.38***	-8.94***	2.66	-11.60***
	3.537	3.854	1.738	3.025	3.561	2.155
Process to license or certify	0.29	-2.67	2.96	-39.72***	-23.56***	-16.16***
	4.096	2.456	2.048	4.916	4.246	1.702
Number of certifications needed	-9.64	-17.85***	8.21***	-15.98***	-3.87	-12.11***
	6.332	6.462	1.635	5.793	6.063	1.891
Overall index of selectivity	-16.41***	-43.22***	26.82***	42.55***	60.83***	-18.28***
	6.645	8.104	2.377	9.044	11.44	3.094
Pay index				38.41***	48.46***	-10.04***
				8.273	7.184	1.775
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:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Standardized Test score									
Recruitment reforms	5.79*** 1.980	12.27*** 4.919	-6.48 4.749	-5.47*** 2.115	11.84*** 4.144	-17.32*** 3.655	-7.11*** 2.200	7.33* 4.365	-14.44*** 3.877
Salary reforms	7.53*** 1.988	1.78 4.022	5.75 3.827	-1.41 2.157	13.22*** 3.389	-14.64*** 2.853	2.11 2.300	13.66*** 3.449	-11.54*** 2.855
Working conditions reforms	-6.48*** 1.998	-4.20 4.277	-2.28 4.081	-20.52*** 1.944	-12.51*** 4.538	-8.01* 4.222	-26.35*** 1.928	-20.63*** 4.819	-5.72 4.512
Retirement reforms	-4.59** 2.137	12.00** 4.670	-16.59*** 4.518	-12.07*** 2.247	8.96** 3.527	-21.03*** 3.093	-14.00*** 2.357	7.25** 3.626	-21.25*** 3.185
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*** [0.290]	0.14 [0.219]	0.18 [0.228]	-2.65*** [0.336]	1.15*** [0.202]	0.84*** [0.209]
Working conditions reforms	-6.89*** [0.384]	-2.08*** [0.353]	-3.92*** [0.338]	-6.37*** [0.380]	0.54* [0.293]	-0.77*** [0.292]
Salary reforms	2.46*** [0.303]	6.70*** [0.421]	7.79*** [0.417]	-1.10*** [0.308]	5.39*** [0.328]	5.69*** [0.328]
Retirement reforms	-2.36*** [0.372]	-0.18 [0.264]	0.67*** [0.254]	-5.18*** [0.439]	0.48** [0.244]	1.03*** [0.236]
Pay index		23.69*** [1.414]			Included	
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*** [5.362]	52.14*** [3.691]	45.30*** [2.698]	38.30*** [2.296]	35.56*** [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*** [3.725]	31.74*** [2.517]	29.77*** [1.802]	27.20*** [1.488]	26.52*** [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* [2.636]	-1.35 [1.829]	0.51 [1.363]	3.23*** [1.143]	4.17*** [1.035]
Working conditions reforms	-11.91*** [2.438]	-13.24*** [1.640]	-12.05*** [1.260]	-12.91*** [1.139]	-13.85*** [1.116]
Pay reforms	5.74** [2.370]	6.54*** [1.639]	6.94*** [1.190]	6.34*** [1.057]	6.78*** [1.019]
Retirement reforms	-6.25*** [2.298]	-9.41*** [1.591]	-10.00*** [1.335]	-8.81*** [1.279]	-8.73*** [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 Induction Program (dummy)	0.11		0.70	0.15			
	[0.305]		[0.456]	[0.357]			
Specific preparation in how to teach reading (dummy)	0.60		0.33	0.50			
	[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)	Country	Year	(1)	(2)	(3)	(4)
Abu Dhabi	1947					Israel	1983				✓
Argentina	1958	✓				Italy	1989			✓	
	1988	✓	✓	✓	✓		1987	✓	✓	✓	✓
	1989		✓				1988	✓			
	1991		✓	✓			1990	✓	✓		
	1993	✓				Kuwait	1960	✓			✓
	2007				✓		1970	✓			
	2016	✓					1981		✓		
Australia	1964	✓					1986		✓	✓	
	1972		✓				1993	✓			
	1987	✓	✓				1998	✓			
	1988			✓		Latvia	2001	✓		✓	
	1991			✓			2004			✓	
	1996	✓					2005			✓	
	2007	✓					2016			✓	
	2010	✓				Lithuania	1997	✓			
	2013	✓					1998	✓		✓	
	2015	✓				Luxembourg	1991	✓	✓		
Austria	1984		✓				1992	✓			
	1988		✓				1996		✓		
	1989	✓			✓		2009	✓			
	1995	✓				Macedonia	2000			✓	
	1996				✓		2008	✓			
	1999		✓	✓	✓		2016	✓			
Azerbaijan	1992	✓	✓	✓	✓	Malta	1990	✓			
	2009	✓	✓	✓	✓		2012	✓			
Belgium (Flemish)	1986		✓			Moldova, Rep. Of	2004				✓
	1991		✓			Morocco	1947				
	1993		✓		✓		1991	✓	✓		✓
	1995		✓				2007	✓			
	1999			✓		Netherlands	1979			✓	
	2003			✓			1984			✓	
Belgium (French)	1986		✓				1985		✓		
	1991		✓				1992	✓		✓	
	1992		✓				1993			✓	✓
	1993		✓				1994	✓	✓	✓	✓
	1995		✓				1995			✓	✓
	1999			✓			1996				✓
	2003			✓			1997			✓	
Belize	1992		✓	✓	✓		1998		✓		
	2010	✓					2000		✓		
Bulgaria	1990	✓	✓				2001				✓
	1996		✓		✓		2002		✓		
	2015	✓			✓		2003		✓		
	2016	✓					2004	✓			
Canada, Alberta	1988	✓	✓		✓	New Zealand	1984				✓

Canada, British Columbia	1986	✓			
	1988	✓			
Canada, Ontario	1947				
	1986	✓	✓		
	1989		✓	✓	✓
	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	✓			
	1987	✓			
	1996	✓			
Northern Ireland	2014	✓			
Norway	1958		✓		
	1971			✓	
	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	✓	✓			1993		✓		
	2005			✓		1995		✓		
	2013	✓				1997		✓		
Georgia	1998	✓				1998		✓		
	2015	✓	✓			2010		✓		
Germany	1991			✓	✓	2013		✓		
	2014		✓							
Greece	1984				✓	Singapore	1984	✓		
	1985	✓					1993		✓	
	1992	✓					1995	✓		
	2010	✓					1999		✓	
	2015				✓	Slovak Republic	1993	✓		
Honduras	2011	✓					2004	✓		
Hong Kong	1947					Slovenia	1994		✓	
							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).