ISER Working Paper Series

Relative Wages and Pupil Performance, evidence from TIMSS

Joshua Fullard
Institute for Social and Economic Research
University of Essex

No. 2021-07
September 2021

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

Economic and Social Research Council

University of Essex

www.iser.essex.ac.uk

Non-technical summary

Just because teachers are intrinsically motivated does not, necessarily, mean that they would not respond to a salary increase in some way that is beneficial to their pupils. This is due to the fact that, in many occupations, we observe that higher salaries often lead to employees working harder, or more productively, in some quantifiable way. One of the most famous examples of this is when Henry Ford (founder of Ford Motor Company) introduced the 'five-dollar day' in 1914 which, in the process of more than doubling wages, resulted in an increase in productivity of up to 70 percent .

While the effect size is fairly modest, which is what we would expect from an intrinsically motivated workforce, we find that teachers do respond to higher salaries and this leads to an improvement in their pupils' test scores. Specifically, over an academic year, a 10 percent increase in teachers' wages has roughly the same the same effect that existing evidence has found for a 1 pupil reduction in class size in Project STAR and found for a one hour increase in weekly instructional time using PISA.

As teachers play an important role in the development of a wide range of their pupils' skills, it is important to understand the role teachers' wages have on other skills developed in school. Indeed, we find that teachers' wages also affect their pupils' well-being, measured by enjoyment of learning.

Using twenty seven years of labour force data we also assess the relative attractiveness of teaching. We find no strong evidence that teachers could earn more in an occupation outside of teaching. Interestingly, we find that teachers who do leave tend to sort into lower, or similarly paying, jobs. This includes teachers with a degree in a STEM subject. This suggests that either teaching is a strong negative signal on the labour market (i.e. the job market doesn't reward teachers) teachers are misinformed about their outside option or teachers who leave the profession are not motivated by money.

Relative Wages and Pupil Performance, evidence from TIMSS

Joshua Fullard†

6th August 2021

Abstract

While it is widely established that higher wages attract more productive individuals into teaching, it is unclear if salaries can be used to motivate existing teachers to work harder, or more productively, in any way that affects pupil outcomes. Using teachers' predicted relative wages, calculated using a novel method of estimating teachers' outside option, we provide evidence that teachers do respond to higher wages and this improves pupil outcomes. Consistent with the predictions of the efficiency wage model a 10% increase in teachers' relative wages improves pupil performance in Science by 0.03sd, Math by 0.024sd as well as their enjoyment of learning by 0.05sd. The magnitude of these effects are similar to a 1 student reduction in class size or an additional hours of weekly tuition.

JEL Classification I12; J31; J45

Key words: Teacher Pay, School Productivity, Efficiency Wage

---

†University of Essex: jjfull@essex.ac.uk

Acknowledgment: This work was generously funded by The Economic and Social Research Council Research Centre on Micro-social Change (grant reference number ES/J500045/1). The author thanks his supervisors Emilia Del Bono, Adeline Delevande and Angus Holford for their guidance and support, my examiners Paul Fisher and Oscar Marcenaro-Gutierrez, and Ludger Woessmann and his team at the ifo Institute who helped shape this paper.

.

1 Introduction

Do teachers' relative wages make a difference to pupil outcomes? This is an important policy question in general, as it is widely established that teachers are the most important school input in the education production function (Chetty et al., 2011, Hanushek 2011a, b, Hanushek et al., 2015, Rivkin et al., 2005). But it is specifically important in the English setting, where the school workforce has faced significant challenges from a decline in quantity (England has faced significant teacher shortages almost continually since the 1940s (Dolton et al., 2003)), to a decline in quality (teachers today are more likely to have lower levels of prior attainment compared to non-teaching graduates (Chevalier and Dolton 2004)).

The literature suggests that there several reasons why teachers' wages might influence pupil outcomes. The first is through occupation choice. When teachers' wages improve, so does the quality of individuals who enter teaching. As teacher quality is the main determinant of school quality (Hanushek 2004), a change in the pecuniary benefits of teaching could impact pupil outcomes through this channel.

Existing evidence suggests that higher salaries improve pupil outcomes by attracting higher quality teachers into the profession. Using a rich administrative dataset linked to pupil test scores Nagler et al., (2015) found that teachers in Florida who joined the profession during a recession (when teaching was relatively more attractive than alternative occupations) were systematically better at raising their pupils' test scores. In the UK this is supported by Nickell and Quintini (2002) who found that the decline in the relative pay of public sector workers in the 1970s and 1980s led to a decline in the quality of men, measured by prior levels of academic attainment, entering teaching.

The second strand of the literature investigates whether wages can be used to motivate existing teachers to work harder, or more productively, in a way that meaningfully affects pupil outcomes. Labour economists have long theorized about how wages can affect labour productivity using efficiency wage models e.g. Shapiro and Stiglitz (1984). An example of how this could occur is through reduced shirking. Effort is costly to the teacher and difficult to monitor, therefore teachers may decide to shirk. But when teachers' wages increase, the outside option becomes less attractive, and the cost of shirking increases. Another possibility is that higher relative wages might improve labour productivity by decreasing the likelihood that an employee has a second job – allowing them to focus on their main job. There is evidence that higher wages decrease the instances of teachers holding a second jobs in

Indonesia (De Ree et al., 2015). However, this is unlikely to be a mechanism in our setting as only 6% of Secondary School and 3% of Primary School teachers have second jobs according to the 2019 Labour Force Survey (LFS).[1]

A final mechanism is that workers respond to an increase in relative wages by improving their productivity due to a fall in perceptions of inequity (Akerlof 1982). According to this hypothesis, when workers feel they are more valued, through a higher relative wage, they work harder. There is suggestive empirical evidence that concerns about fairness and equity do influence effort, see Fehr et al., (2009) for a review of this literature. Therefore, teachers' higher relative wages could drive the productivity of teachers, and thus pupil outcomes, through the mechanism of feeling more valued.

Theoretical and qualitative studies suggest that salary increases are an important mechanism for motivating and encouraging teachers to work harder (Hanushek et al., 1999, Webb and Valencia 2005). However, other empirical evidence suggests that an unconditional salary increase has no effect on pupils' performance. Most famously De Ree et al., (2015), using data from a randomized experiment in Indonesia, found that doubling teachers' pay had no meaningful effect on students' learning, although it did reduce the likelihood of a teacher holding a second job and improved job satisfaction. Although there is some evidence in the UK that pupils perform better when a teacher's outside option is lower (Britton and Propper 2016), the majority of the literature finds no correlation between changes in teachers' salaries and student outcomes (Hanushek 1986). While the existing evidence suggests that an unconditional pay rise does not impact pupil performance, there is strong evidence that teachers respond positively to performance-related pay in a variety of settings around the world (Atkinson et al., 2009, Kingdon and Teal 2007, Loyalka et al., 2019, Woessmann 2011, Zhang et al., 2019).

An important challenge in all these studies is identifying what teachers' relative wages actually are. In this paper, we use twenty seven years of the Labour Force Survey (LFS) to identify teachers' relative wages using a novel method of estimating teachers' outside option, which takes into account differences in job security, and that entry into teaching is a choice. In doing this we demonstrate that, when we account for non-random selection and differences

[1] However, using the LFS we are unable to identify if these second jobs are during term time, or not. Given that teachers, during term time, typically work a 52hr week and are 12% more likely to be dissatisfied with their working hours, compared to the average graduate, it is most likely that the majority of these second jobs are taken out of term time and, therefore, do not impact teacher productivity (Dolton 2004).

in job security, teachers' salaries compare favourably to their outside option. One of the main contributions of this paper is that we demonstrate that failing to account for the relative job security of teaching underestimates teachers' relative wages. While the effect in our context is modest (5% for young graduates (under 30's) and between 1 to 2% for older graduates (30 or over)) failing to account for job security could have a large effect on the relative wage estimates for teachers in other settings, such as Spain, where the graduate unemployment rate tends to be higher.

Using the relative wage estimates from the LFS we impute these to five waves of the Trends in International Mathematics and Science Study (TIMSS). Then estimate the effect of relative wages on pupils' test scores and enjoyment of learning by regressing pupil outcomes on teachers' predicted relative wages controlling for a rich set of classroom, school and household level covariates. The effect on pupils' test scores is relatively small, with a 10% increase in teachers' salaries leading to a 0.03sd improvement in test scores, which is similar to the benefit of an additional hour of weekly tuition (Lavy 2015). We also find that teachers' relative wages lead to an increase in their pupils' well-being, measured by enjoyment of learning.

We contribute to the literature on teachers' wages and pupil outcomes in the following ways: first we derive a measure of teachers' relative wages that accounts for differences in job security. This is an important contribution as existing evidence shows that job security plays an important role in the decision to become a teacher and a failure to include this underestimates the returns to teaching (Heinz 2015, Priyadharshini and Robinson-Pant 2003). Second we use a rich data set that allows us to estimate the effect on test scores (Mathematics and Science) and pupil well-being, measured by enjoyment of learning. The existing literature has exclusively focused on the effect of teachers' wages on test scores and other measures of cognitive performance.[2] As teachers play an important role in the development of a wide range of skills, it is important to understand the role teachers' wages have on other skills developed in school.

The empirical analysis is set in England. This is an important policy setting as the government is currently undergoing a wide range of sweeping policy reforms. The most prominent of which is the commitment to increasing teachers' initial wages to £30,000, an

[2] This is an important finding because empirical evidence from our setting shows that literacy and numeracy skills have a high value in the UK labour market even when we control for education (Vignoles et al., 2011).

increase of 24%, to attract the highest-achieving graduates into teaching.[3] While making teaching among the highest paid graduate occupations is likely to improve the quality of graduates entering the profession it will take time for new teachers to be recruited, trained and integrate into the education system. This paper shows that policymakers should expect benefits from raising the salaries of existing teachers. This paper is organised as follows; in Section 2 we introduce three methods of estimating teachers outside option and consider if teachers in England are underpaid, in Section 3 we introduce the data on pupil outcomes and the empirical strategy, in Section 4 we present our main results, in Section 5 we present our robustness checks and in Section 6 we discuss our results and conclude.

2 Relative Wages

The majority of the literature that investigates the effects that teachers' wages have on pupil outcomes has exploited differences in teachers' wages relative to occupations outside of teaching. This is because using the changes in teachers' absolute wage requires us to assume that all other factors that affect behaviour, such as wages in an alternative occupation, are held constant (Sharir and Weiss 1974). In many settings this assumption does not hold, therefore, many papers that exploit absolute wage differences do not make causal claims. For example, using cross-sectional data Dolton and Marcenaro-Gutierrez (2011) find that countries that pay their teachers higher salaries tend to perform better on international tests. However, there are settings where using absolute wages is appropriate for causal inference. The first is under a large policy intervention where changes in the outside option is likely to be inconsequential. Such a setting includes Indonesia when teachers' salaries were doubled between 2006 and 2015 (De Ree et al., 2015). The second is exploiting wage variations in a setting where non-teachers wages are plausibly similar. As we are not exploiting a significant policy intervention and non-teachers' wages vary in our setting we will exploit variation in teachers' relative wages.

To investigate whether teachers respond to a change in their relative wage in a way that affects their pupils' test scores or well-being, we must first estimate their outside option. The most common measure of estimating how much teachers would have earned had they not gone into teaching is by comparing the earnings of current teachers to the earnings of some non-teaching group. Traditionally, the comparison group used was workers in non-manual

[3] See Fullard and Zuccollo (2021) for a summary of the latest policies related to the supply of teachers in England and Fullard (2021b) for a discussion of the Department for Education's current pay policy in England.

occupations (Nickell and Quintini 2002). Although this data is both easily accessible and makes sense in a historic context, these groups are sufficiently different, such that any difference in earnings is likely to be due to differences in workers' characteristics. For this reason, using non-manual occupations is not a sensible approach in our context as it is unlikely to capture teachers' outside option. A comparison group that makes more sense in our context is university graduates. All teachers in England are generally required to have a university degree, meaning that all the occupations available to graduates are, in principle, also available to teachers (Hermann and Diallo 2017). This will be our first measure of teachers' outside option. This will be referred to as "Non-Teachers' Wages (Normal)".

Using this measure consistently finds that teachers earn less than the average graduate.[4] Although this might explain why policymakers struggle to recruit graduates from the higher end of the ability distribution, this does not necessarily mean that teachers could earn more in their outside option. This is due to the fact that selection into teaching is non-random even among those with a university degree. Therefore, the average graduates' earnings are unlikely to reflect the salary that teachers would earn if they left teaching.

To get around non-random selection, Chevalier et al., (2007) used a matching strategy to estimate teachers' outside option by comparing teachers' with non-teachers' who looked most like them based on observable characteristics. Using this approach will be our second measure of teachers outside option. This will be referred to as "Non-Teachers' Wages (Matched)".

Using this strategy, Chevalier et al., (2007) they find no evidence to suggest that teachers are underpaid. While these conditional estimates are more likely to reflect teachers outside option it fails to account for another significant benefit of teaching – job security. Existing research shows that job security plays a significant role in the decision to go into teaching, failing to account for this may further underestimate the pecuniary benefits of teaching (Priyadharshini and Robinson-Pant 2003). To estimate teachers' outside option we will modify the matching strategy used by Chevalier et al., (2007) to take this into account. This will be our third measure of teachers outside option and will be referred to as the "Labour Market Returns to Teaching".

[4] The Department for Educations 2019 report uses this method where they state that the earnings gaps between teachers and other graduate professions are "important contributory factors in the recruitment and retention problems facing the teaching profession".

We recognize that our matched estimates may still be affected by differences in teaching and non-teaching graduates' unobserved characteristics. One way to get around the difference in unobservable characteristics is to compare the earnings of current teachers to those of former teachers. By doing so, Scafidi et al., (2006) shows that very few teachers who leave teaching enter better-paid occupations. However, this does not tell us how much the average teacher would earn if they quit as attrition is non-random. In the supplementary material, we also estimate teachers outside option by using a matching strategy to compare the wages of current teachers to the earnings of former teachers who look most like them based on observables.[5]

2.1 How we estimate teachers' relative wage

Our first method of estimating teachers' relative wages is to compare the average teachers' wage to the average graduates' wage. We use 26 years of the Labour Force Survey (1993 – 2019) and restrict our sample to those who are: university graduates, working age, full time employed and report weekly earnings greater than the expected minimum wage. The average teacher wage, $w_{Teacher}$, is the average weekly wage of all individuals whose main occupation is teaching and who are currently teaching. The average graduate wage is the average wage of non-teaching graduates',$w_{Non-Teacher(Normal)}$. All wages are CPI adjusted to 2019 prices. The difference in the natural log of teachers' and non-teaching graduates' earnings is our first measure of teachers' relative wages. This will be referred to as "Wage Difference (Normal)", as shown in equation 1.

$$\textit{Wage Difference (Normal)} = Ln\, w_{Teacher} - Ln\, w_{Non-Teacher(Normal)} \qquad (1)$$

Table 1 shows the teachers' wage (column a) and graduates' wage (column b) for each year between 1993 and 2019. Although the difference in teachers' and non-teaching graduates' wage might partially explain why policymakers struggle to attract the highest-achieving graduates into teaching, it does not mean that teachers currently face a pay penalty. This is because the composition of individuals who enter teaching have different characteristics to those who do not. For example, using the Destination of Leavers from Higher Education (DLHE), we observe that between 2003 and 2012, 88% of the graduates on initial teacher

[5] Although we will discuss the earnings of teachers who quit teaching we do not necessarily have enough power to use former teachers and our matching strategy so we do not discuss this in detail in the main text but the data is available in the additional material.

training programmes were female (vs 56% of graduates not enrolled), 95% were white (vs 85%), 96% state school educated (vs 86%), with only 0.4% from Oxbridge (vs 4%) and 1.7% from Russell Group (vs 26%) institutions (Fullard 2018).

To account for the differences in observable characteristics, we follow Chevalier et al. (2007) and estimate teachers' outside option by using propensity score matching (PSM). This is our second measure of teachers outside option, $w_{Non-Teacher(Matched)}$. Using this we construct our second measure of teachers' relative wages – "Wage Difference (Matched)", as shown in equation 2.

$$\textit{Wage Difference (Matched):}\ \ Ln\, w_{Teacher} - Ln\, w_{Non-Teacher(Matched)} \qquad (2)$$

PSM is a method first proposed by Rosenbaum and Rubin (1983) designed to balance the distribution of baseline covariates between a treatment group (teachers in our case) and a control group (graduates). This strategy allows us to estimate the treatment effect on some outcomes (wages) by comparing the treatment group to members of the control group who look most similar to them on observable characteristics. This is achieved by first estimating the conditional probability of an individual receiving the treatment (i.e. becoming a teacher) given observable characteristics. We do this by regressing observable characteristics on the treatment status using a logistic regression.[6] Then we assign each member of the treatment group to their nearest neighbour in the control group based on their probability of receiving the treatment (propensity score). Within these pairs we use the outcome of the individual in the control group to estimate the counterfactual of the treatment group. Appendix Table 1 shows that teachers are less likely to be men than the graduate population (42% vs 65% in 2000 and 37% vs 59% in 2010) but when we use the matched sample, the difference falls (from 65% to 45% and 59% to 38% respectively). This highlights the importance of controlling for differences in observable characteristics.

For our estimates of teachers' outside option to be unbiased we must have common support (Heckman et al., 1997). Specifically we must compare individuals in the treatment group to individuals who look similar to them in the control group. To test this condition we perform the minima and maxima comparison (Caliendo and Kopeinig 2008) by dropping all observations that have a propensity score which lie outside the minimum and maximum of

[6] The observable characteristics we control for are sex, age, ethnicity, home region and degree subject.

either the treatment or control group. This has no effect on our estimates. An additional problem with common support can occur if the density in the tails of the distribution is very thin. To test this, we follow Lechner (2004) and do a sensitivity check by replacing the minima and maxima with the tenth smallest and tenth largest observation. Doing this also has no significant effect on our estimates. Therefore, we are confident that our matched estimates are not affected by problems related to common support.

We use PSM to identify the conditional difference in teachers' and non-teachers' salaries because it is simple to estimate, it does not rely on exclusion restriction or functional form to control for differences between teachers and non-teaching graduates, and it is easy to check if covariates are balanced, as shown in Appendix Table 1 (Williamson et al., 2012). We use two alternative strategies (inverse probability weighting (IPW) and regression adjustment (RA)) to estimate teachers outside option as a robustness check. These are presented in figure 1 in the appendix and show that while there are some differences prior to 2000 (when the sample of teachers was roughly 800 each year) from 2001 the estimates are largely similar (roughly 1,500 teachers each year).[7]

Although improved, our second measure does not account for the fact that teaching, as an occupation, has significantly lower unemployment levels. Given that job security plays an important role in attracting graduates into teaching, failing to account for this benefit underestimates the returns to teaching. We account for this benefit by weighting the teacher and non-teacher wage estimates with a teacher and non-teacher unemployment rate obtained using the LFS. This is our final measure of teachers' relative wages – "Labour Market Returns to Teaching". We estimate this separately using $w_{Non-Teacher(Normal)}$ and $w_{Non-Teacher(Match)}$ in equations 3a and 3b respectively.

Labour Market Returns to Teaching (Normal) = (3a)

$$P(Emp|Teacher, Age, Sex)Ln\, w_{Teacher} + \left(1 - P(Emp|Teacher, Age, Sex)\right)Ln\, w_{JSA} - P(Emp|Non-Teacher, Age, Sex)Ln\, w_{Non-Teacher(Normal)} + (1 - P(Emp|Non-Teacher, Age, Sex))Ln\, w_{JSA}$$

[7] IPW is known to perform better when sample sizes are smaller. Our main results are robust to using teachers relative wages estimated via IPW, RA or alternative propensity score matching strategies (i.e. kernel).

Labour Market Returns to Teaching (Matched) = (3b)

$$P(Emp|Teacher, Age, Sex)Ln\, w_{Teacher} + \big(1 - P(Emp|Teacher, Age, Sex)\big)Ln\, w_{JSA} - P(Emp|Non-Teacher, Age, Sex)Ln\, w_{Non-Teacher(Match)} + (1 - P(Emp|Non-Teacher, Age, Sex))Ln\, w_{JSA}$$

Where w_{JSA} is the unemployment benefits they would be eligible for.

The teacher unemployment rate is the sum of unemployed individuals whose last job was teaching divided by the number of teachers plus the quantity of unemployed teachers. While this measure does miss those young people who are unable to find their first teaching job, using the alternative (e.g. individuals who are qualified to teach) would not be any better. A high proportion of young people who finish teacher training decide not to go into teaching (1 in 5 men and 1 in 10 women) and this is driven by preferences and not an inability to find a job (Each year roughly 3,000 more teachers leave the profession than enrol onto teacher training programmes).[8] Using the annual statistics from the Department of Work and Pensions (DWP), we calculate the cost of unemployment by estimating the unemployment benefits (Job Seekers Allowance (JSA)) that everyone would be entitled to given the year, their age and sex – similar to the wages we adjust all expected benefit entitlements to 2019 prices using the CPI.

In the following section we will show how teachers' wages compare to non-teaching graduates wages and how accounting for differences in observable characteristics and differences in job security affect our measure of teachers' relative wage.

2.2 Comparison across different measures of teachers' relative wages

Comparing the earnings of teachers to non-teaching graduates we find that from 1993 to 2019, the average teacher earns around 13% less than the average graduate (Table 1 column b and figure 1). The difference in pay was largest in the late 1990s but fell to under 10% prior to the 2010 public sector pay freeze. But since then the difference has risen to 14%.

A particularly striking observation is that young teachers' wages are highly competitive and remain this way despite the public sector pay freeze (Figure 2 LHS black solid line vs black

[8] Specific details about how we use the LFS to calculate teachers' and non-teachers unemployment rate is available in the supplementary material.

dashed line). However teachers' wages do grow at a significantly slower rate than non-teaching graduates wages over the age distribution - teachers in their 30s, 40s and 50s earn around 20%, 23% and 15% less than the average graduate in their respective cohorts (Figure 2 the wedge between the solid and dashed lines grow over the age distribution). This suggests that young people who quit teaching due to pecuniary reasons are motivated by expected future earnings and not current earnings.

Using matching to account for non-random selection, we observe that the average difference in teachers' pay falls from 13% to 3% (table 1 column c). Although there is still evidence that, during the 1990s and after the public sector pay freeze, teachers were paid less than their outside option the magnitude falls significantly (to 9% and 5% respectively). Additionally, the 2019 data suggests that teachers do not, currently, face a wage penalty. However, this may, in part, be due to changes in the composition of the workforce. Teachers' real wages have fallen since 2010 which may have led to the teachers who face a larger pay penalty leaving the occupation at a higher rate – thus changing the composition on both observable and unobservable characteristics. Indeed the proportion of male teachers has fallen (37% in 2010 to 34% in 2018) as has the proportion of teachers with a degree in Mathematical Sciences (14% vs 10%) or Biological Sciences (7% vs 5%). See Fullard (2020b) for a summary for a summary of the trends in the diversity of the school workforce in England since the 2010 public sector pay freeze.

Accounting for the difference in job security, using our final method has a fairly modest effect (making teaching 1 to 2% more attractive) on our estimates for any group over the age of thirty as older graduates have a very low unemployment rate (under 3% between 1993 and 2019 vs 1.7% for teachers). However, young graduates have a higher unemployment rate (e.g. 5% between 2013 and 2016) and taking this into account does make teaching up to 5% more attractive. The job security young teachers enjoy combined with their relatively high earnings reinforces the notion that young people typically have a significant pecuniary benefit to enter, and remain in, the profession.[9][10]

[9] These figures are not reported.

[10] Although matching accounts for differences in observable characteristics teaching is a vocational occupation. Therefore, these estimates are likely to be biased due to differences in unobservable characteristics. Comparing the earnings of current teachers to the earnings of former teachers, we find no evidence that those who quit teaching entered higher paid occupations between 1993 and 2010. However since the public sector pay freeze, we find that teachers who left the occupation, typically enter occupations that pay up to 9% more than teaching. But this does not mean that current teachers could earn as much as individuals with the highest outside option, ceteris paribus, are more likely to leave e.g. Friedman and Kuznets (1945).

Due to its policy relevance we will briefly discuss how teachers' relative wages differ by school phrase (Primary vs Secondary) and educational background (STEM vs Non-STEM) in the following sections.

2.3 Primary and Secondary School Teachers

In this paper, we combine all teachers together (secondary, primary and nursery or special education) so that we can achieve: i) a sample size sufficient to estimate the relative wages by sex and age, and ii) intertemporal consistency – prior to 2001 the LFS does not allow us to identify which type of teacher the respondent is.[11] However, it is still interesting to look at the differences between different categories of teachers (these figures are not reported). For example, comparing the earnings of secondary (primary) school teachers to the earnings of non-teaching graduates between 2001 and 2019, we find that teachers earn between 5-12% (13-23%) less. Although primary and secondary school teachers are on the same national pay scales, it is unsurprising that primary school teachers earn less than secondary school teachers, relative to the average graduate, due to differences in the workforce composition. Teachers' wages are linked to experience and primary school teachers tend to be significantly less experienced (according to the 2018 School Workforce Census 33% (24%) of classroom primary (secondary) school teachers are under 30 while 13% (16%) are over the age of 50).

Using matching to account for non-random selection we find that, prior to the public sector pay freeze, both primary and secondary school teachers' wages were fairly similar to their outside option. While both suffered significant pay penalties due to the pay freeze (up to 8% for secondary and 11% for primary) changes in the composition of the school workforce mean that there is no strong evidence that secondary school teachers face a pay penalty today (the secondary school teachers with the highest outside option left the profession) but the average primary school teacher does face a pay penalty of around 8% today (2019).

2.4 Relative Wages of STEM and Non-STEM Teachers

In England, teacher recruitment and retention challenges are more severe in areas that require a degree in a STEM subject (i.e. Science, Technology, Engineering and Mathematics). Given

[11] From 1993-2000, the LFS's main occupation code does not allow us to identify the type of teaching professional the respondent is.

that STEM graduates typically earn more in non-teaching jobs, differences in relative wages could explain this.[12]

Table 2 shows that teachers with a university degree in a STEM subject typically face a wage penalty for entering the teaching profession. However, the magnitude of the penalty has fallen dramatically from over 12% in the mid-90s to around 6% (column a). This suggests that teaching has become more attractive to STEM graduates despite the public sector pay freeze. We also observe that non-STEM graduates are relatively better off in teaching as they typically earn as much in teaching as they would in an outside option, if not more (column b).

Given that STEM teachers have a higher outside option we would expect STEM teachers who leave teaching to enter higher-paid occupations, on average. But we do not have any strong evidence that this is the case (column c). One possible reason for this might be that the skills a teacher acquires are so occupation-specific that they constrain future labour market opportunities. However, we also observe that, since the public sector pay freeze, non-STEM graduates who leave teaching appear to be entering higher paying occupations (10% higher since 2015). While it is possible that teaching might constrain future labour market opportunities differently for STEM and non-STEM graduates, it is possible these graduates also have systematically different preferences in the types of jobs they would be interested in outside of teaching.

3 Teacher Pay and Pupil Outcomes

Having derived relative wage measures, we will now estimate the effect of these measures on pupil performance using measures of pupil outcomes from five waves of the Trends in International Mathematics and Science Study (1995, 2003, 2007, 2011 and 2015). Specifically we are interested in pupils' test scores measured by performance in Science and Maths achievement tests and a measure of well-being, here represented by students' self-reported enjoyment of learning.

3.1 Empirical Strategy

To estimate the effect on pupil performance (enjoyment of learning), we will perform a least-squares regression of test scores (learning preferences) on relative wages controlling for a set

[12] To get a sample that is large enough to estimate teachers' relative wage by degree subject, we combine the two preceding and two following LFS years. For example, for the STEM and Non-STEM wages in 1995 we merge the LFS years 1993-96. Further details are available in the supplementary material.

of pupil, class and teacher characteristics. Using test-score (student survey) data from different grades (4 and 8) and subjects (Math and Science), we estimate the following:

$$\mathrm{Y}_{ict} = \beta_0 + \beta_1 wage_{ct} + \beta_2 \boldsymbol{X}_{ict} + \theta_t' + \varepsilon_{ict} \quad (4)$$

Where Y_{ict} is the test score of students i in class c in year t. The test scores are originally standardized so they have an international mean of 500 and a standard deviation of 100. As we are not using the international dataset, we re-standardize the scores within our sample of English students to have a mean of 0 and a standard deviation of 1 for ease of interpretation. To estimate the effect of relative wages on non-cognitive skills we replace Y_{ict} with a dummy that indicates whether the student i in class c in year t enjoys learning, or not.

Our regressor of interest, $wage_{ct}$,is the difference in the natural log of teachers' and the natural log of non-teachers' wages of the teacher in class c at time t . Where the differences are either the simple difference in earnings or the weighted difference shown in equations 1, 2 and 3 and non-teachers' earnings are estimated using either the average graduates' earnings or matching. $\boldsymbol{X}$ is a vector of controls for pupil, class and teacher background characteristics. This vector includes the relative student age measured in the difference in months from the median, the students' sex measured as a male dummy, the size of class above the median (by subject). To control for the child's socioeconomic status, we use five dummies to control for the number of books at home (0-10, 11-25, 26-100, 101-200, 200+) a dummy if they have a computer at home and a dummy if they speak English at home. We also control for teacher characteristics, these are: sex, experience (using 5 dummies), and age (using 6 dummies for different age groups.). The last term, θ_t represents year fixed effects. Our coefficient β_1 is our parameter to be estimated. ε_{ict} is our pupil specific error term observed at time t in class c. Our standard errors are clustered at the classroom level because the unobservable component of pupil outcomes in the same class is likely to be correlated (e.g. class resources, time spent on certain topics) and because predicted teachers' pay is constant within classrooms.

The difficulty of interpreting β_1 as a causal effect, in equation 4, is that the variation in teachers' relative wages may not be exogenous to the variation in pupil performance. Indeed there are two forms of selection that could bias our results. The first of these is between school selection, in which students from more affluent households or higher ability, could select into schools that put a lot of emphasis on academic achievement and pay their teachers higher salaries (upward bias). Conversely, we might have situations where schools which have a higher proportion of students from less affluent backgrounds, or lower academic ability, might have to pay a wage premium to attract teachers (downward bias). The second is within school selection, in which more able students might be separated into different classes and taught by more able/higher paid teachers.

Between school selection is potentially an issue in our setting: while teachers' pay scales are determined at the national level, schools have the freedom to pay teachers any amount within the centrally defined minimum and maximum, for a given level of experience. We think that within school selection could also be an issue for the older (grade 8) students in our sample, as most schools in England tend to sort students into classes by ability during secondary school. Whatever the source of endogeneity, it is possible that variation in teachers' wages, $wage_{ct}$, is associated with variation in pupil outcomes, Y_{ict} due to these other reasons and not simply because it affects teachers' productivity. Therefore, using actual teachers' pay would not provide us with a causal effect of teachers' wage on pupil performance.

In the TIMSS data, we do not observe actual teacher wages for each class, i.e. $wage_{ct}$ and are therefore unable to estimate equation 1. Instead, we use the LFS data to obtain a measure of teachers' wages as predicted by a model where we use age, sex and year as explanatory variables. Using these variables, we then impute the estimated wages to the TIMSS data. This way our wage measure changes by class only to the extent that classes are taught by teachers of a different sex and age. Ultimately, what we are exploiting is simply variation in teachers' wages by year, sex and age. Consequently, β_1 is less likely to be affected by a problem of endogeneity and could be interpreted as the causal effect of teachers' relative wages on pupil performance and enjoyment of learning.

Since $wage_{ct}$ is an estimated regressor – relative wages are imputed from the LFS and assigned to teachers based on the teachers' sex, age and the year they are observed - standard errors calculated in the usual way are biased. This is due to the fact that teachers' predicted relative wages has additional sampling variance that needs to be taken into account when we calculate the variance of our final parameter estimates. To obtain unbiased standard errors, we follow Chevalier et al., (2007) and bootstrap the estimates (500 times).

As a robustness check, we exploit variation within schools with a similar level of attainment by using school attainment fixed effects to show that our main results are robust to this more conservative specification. We do not use school attainment fixed effects in our main model because the schools' prior attainment data is not available in the most recent wave (2015) and therefore including this forces us to drop roughly 20% of our sample.

3.2 Data

The TIMSS data comes from tests in Science and Mathematics that are administered by the International Association for the Evaluation of Educational Achievement to nationally representative pupils in grades 4 (approximately age 9) and 8 (age 13). TIMSS is an international assessment designed to assess and compare the achievements of young people in more than 60 countries. Along with the tests, TIMSS also contains a rich amount of data on the students, the schools they attend and the teachers who teach them. We merge the pupil performance data with the pupil and teacher surveys together from the 1995, 2003, 2007, 2011 and 2015 TIMSS surveys which gives us our data set.

The TIMSS 4^{th} Grade assessment in England is taken by pupils in Year 5 (primary school) and the 8^{th} Grade assessment is taken by Year 9 pupils (secondary school) as long as the average class age is over 9.5 (13.5) years old at the time of assessment for Grade 4 (Grade 8). However, the 1995 and 2003 TIMSS waves were assigned based on age and not years of schooling. This means that the Grade 8 tests were taken by students in two adjacent grades that contain the largest proportion of 13 year olds (or 10 year olds for Grade 4). In England, this means that the grade 8 tests were taken by Year 8 and Year 9 pupils and the Grade 4 tests by Year 4 and Year 5 pupils. As a consequence the average ages of pupils are moderately lower in these waves.

TIMSS is designed to be nationally representative of pupils. The assessment is randomly assigned to classes using a stratified two-staged cluster sample design. First schools are sampled with probabilities according to their size from the list of all schools in the population that contain eligible students. They are stratified according to demographic characteristics, but the exact variables used differ by country. The most common are: region, urbanization and socioeconomic indicators. The second stage is selecting one or more classes from those eligible within the selected school. Pupils with additional educational needs who are unable to follow the test instructions are excluded, as are students who have received less than one year of instruction in the language of the test. But students who have low prior attainment and/or behaviour problems are eligible to participate. Roughly 2% of children are excluded from the sample in England for one of the reasons above. Conditional on selection and eligibility, participation rates in England are high (96%).

In this paper we drop all pupils where we either cannot match the pupil to a teacher, or where the age and/or sex of the teacher who taught them is missing. We drop these students because we assign teachers' relative wages based on their sex and age –if these are missing, we are unable to assign them a teaching and non-teaching wage. In addition, we drop cases where the student did not complete the home questionnaire, or those who did not complete the questions we use to control for SES. This is because a student's socioeconomic status is an important predictor of cognitive performance.[13] Across the 5 waves we drop 3,245 students in Grade 4 and 4,225 (9,514) students in Grade 8 Math (Science). This leaves us with a sample of 25,346 Grade 4 pupils in both Maths and Science and 15,177 Grade 8 pupils in Math and 17,302 in Science. Table 3 shows that the young people who we drop from our analysis achieve lower scores on the Mathematics and Science assessment, report a lower enjoyment of learning and tend to be marginally younger.

3.3 Assigning Teachers' Relative Wages

We assign each teacher in TIMSS a teaching and non-teaching wage based on their age, sex and the year they are observed. Our wage estimates are obtained from the LFS (see section 2) by combining the two preceding and two following LFS years to each TIMSS year. For example, we merge the LFS years 1993-1996 and use this sample to estimate the relative wages of teachers observed in the 1995 TIMSS wave (see supplementary material). We assign each teacher the following: a teacher wage, a non-teacher wage estimated using matching and a non-teacher wage estimated not using matching. We also assign each teacher a teacher unemployment rate and a non-teacher (graduate) unemployment rate based on their sex, age and year observed using the LFS. Finally, each teacher is assigned an estimate of the unemployment benefit entitlement (JSA) by applying Department of Work and Pensions (DWP) rules (for a give age, sex and year the JSA entitlement is the same for both teachers and non-teachers).[14]

[13] Across the 5 waves only 126 young people who completed the home questionnaire did not complete the questions we use to control for SES. Including these young people in our model using a missing dummy has no impact on our results.
[14] See the DWP website: https://www.gov.uk/benefits-calculators and https://www.gov.uk/government/collections/dwp-statistical-summaries

Using this information we compute each teacher's difference in wages – the natural log of their predicted teacher wage minus the natural log of their predicted non-teacher wage. We do this twice, first using non-teachers' wages estimated using matching and second using the outside option estimated using the average graduate's wages.

Finally, we account for both the differences in job security and the cost of unemployment. It is important to note that all the wages are logged so that the results show the effect of a one percent change in wages or relative wages on pupil performanceTable 4 shows the means and standard deviations of these different measures. From this table rows 1, 3 and 5 show that the average graduate earns more than the average teacher but when we account for non-random selection, there is no evidence that the teachers in our sample, on average, face a pay penalty (rows 2 and 4). Additionally, teachers are significantly less likely to be unemployed than graduates (1.7% vs 3.1% for Grade 4 teachers and 1.8% vs 2.9% for Grade 8 teachers). Therefore, when we combine these differences we find that the teachers in our TIMSS Grade 4 and Grade 8 samples do not, on average, face a pecuniary penalty for remaining in the profession.

3.4 Descriptive Statistics

Table 5 presents the descriptive statistics for the students who took the TIMSS assessment. Consistent with the design of the assessment, where the average class age for grade 4 (grade 8) had to be higher than 9.5 (13.5), the grade 4 students are typically 10 years old and grade 8 students are 14 years old. There is an equal gender split for both grades.

More grade 8 students live in a household with a home computer (94% vs 87%). Grade 4 pupils are more likely to be taught by a teacher with less than 4 years' experience (27% vs 22.5% for grade 8 Math and 21.4% for Science). The younger pupils are also more likely to be taught by a teacher 25 or under (7.8% vs 6.1% for grade 8 Math and 4.9% for Science) and over 60 (15.4% vs 2.5% Math and 2.1% Science). Consistent with the gender gap in primary teaching the young pupils are much less likely to be taught by a male teacher (26%) than the older pupils where it is a relatively even gender split.

Table 6 presents the relationship between our controls and outcomes in a multivariate regression that does not include our regressor of interest. First looking at the differences in school and student characteristics, the first row shows that pupils in larger classes tend to do better. Consistent with the literature this suggests that there is non-random sorting in England where pupils who need more individual attention tend to be sorted into smaller classes (Woessmann and West 2002). Similar to the literature, the second row shows that, within cohorts, older students perform better in Math and Science – an increase in age by one month is associated with an increase in pupil performance by 0.03sd for Grade 4 and 0.01sd for Grade 8 (Bedard and Dhuey 2006, Strøm 2004). Consistent with the existing evidence of gender gaps the third row shows that male students tend to outperform female pupils in both Maths and Science and the gap gets larger with age (Contini et al., 2017, Muñoz 2018).

There is a large body of existing literature that demonstrates the strong relationship between socioeconomic status and academic achievement; these include Duncan and Murnane (2011) and Dahl and Lochner (2012). As we do not know parental income, occupation, or highest educational attainment we use two different controls for SES (Rows 4-9). Our first proxy for SES is books at home. Rows 4-8 show that pupil achievement increases with the quantity of books in the home and the achievement gap is steady for both grade 4 and grade 8 pupils. Consistent with Hanushek et al., (2019), who found that the achievement gap has remained fairly constant between 1954 and 2001 in the US, figure 4 shows that the disadvantage gap in Math has remained fairly constant over the last two decades in the UK. But the difference in Science achievement between the most advantaged pupils and the least advantaged pupils fell by 0.4sd. Our second proxy for SES is having a computer in the home, which (as shown in row 9) has no effect on pupil performance.

Row 10 shows that there is a positive relationship between speaking English in the home for Science performance while there is a negative relationship with grade 8 Math performance, this is consistent with existing evidence in England that uses TIMSS (Greany et al., 2016).

The literature on teacher effects consistently shows that teachers have a significant impact on pupil performance. Among the characteristics which are considered important include the teachers' sex, years of teaching experience and age. We do not observe any aggregate effects of teachers' sex on pupil performance apart from Grade 8 Science, in line with the existing literature we also observe that the pupils with the least experienced teachers and the youngest teachers tend to perform worse (rows 12 – 23).

4 Estimation Results

The literature has predominately focused on the effect of relative wages on pupil performance therefore we will introduce these results first (Table 7) and then present the results on learning enjoyment (Table 8).

In our data we only observe one teacher for each student. For the young students (grade 4), this is their only teacher. For the older students (grade 8), this is one of many teachers, likely to be of a diverse profile.[15] As a consequence spill-over effects or complementarities could attenuate any wage effects we find for the older students. For example, the benefits that a pupil who is taught by a more effective Science teacher, who is more motivated due to a higher relative wage, might make a positive difference to their Maths score, and vice versa (spill-over effect). Alternatively, having a more effective maths teacher might increase the returns of having a more effective Science teacher (e.g. by improving numeracy skills). As there is evidence that these effects exist in one form or another it will be fairly difficult to identify a wage effect on secondary school pupils (Bryson and Papps 2016, Kinsler 2016, Sun et al., 2017). Therefore our main focus will be on the results of the primary school pupils. Our estimates for grade 8 pupils are smaller and less precise than our estimates for the younger pupils, which is consistent with spill-over effects, but we cannot assess their magnitude. The results for our secondary school pupils are available in the appendix (Tables 2 - 4).

4.1 Teachers and non-Teachers Wages

Column 1 in table 7 shows that, consistent with an efficiency wage model, the effect of teachers' wages on pupil performance in grade 4 Science is positive. An increase in teachers' wages by 10%, which is roughly how much teachers would expect their salaries to increase after acquiring an additional year of experience (for example moving up from the lowest pay band (M1 to M2) on the 2019-20 pay scales), improves pupil performance by 0.024sd. The effect of such an increase in wages is similar to that identified in the literature from a 1 pupil reduction in class size (Krueger (1999) 0.03sd) and a 15% decrease in traffic pollution Heissel et al., (2019) 0.024sd). What these estimates mean in a wider policy context will be discussed in detail in section 6. The effects on Grade 4 Math performance, columns 8 – 10, display a similar pattern although the magnitude is smaller.

[15] In a scenario where students are taught by equally effective teachers with correlated characteristics (and therefore are estimated to face the same relative wage) this would not be a problem. However, this is unlikely to hold as secondary school teachers are more diverse than primary school teachers (i.e. 50% male teachers in secondary schools vs 26% in primary).

4.2 The Difference in Relative wages

In the previous section we observe that teachers' wages are positively associated with pupil performance and non-teachers' wages are negatively associated with pupil performance. Therefore, when we take the difference in teachers' and non-teachers' wages we would expect to observe a positive relationship.

Regressing pupils' Science performance on the difference in teachers' and our matched non-teachers' wages (table 7 column 4), we find that a 10% increase in teachers' relative wage causes a 0.0265sd increase in pupil attainment, statistically significant at the 1% level. This effect is stronger than using the non-matched outside option (0.0208sd column 5). While this does provide some evidence that our matched estimate might be a better measure of teachers' outside option, than the average graduates' wage, the two estimates are statistically indistinguishable. We observe a similar effect on Math performance, but with a smaller effect size (column 11 – 12).

Our relative wages' estimates are similar to Britton and Propper (2016), in which a 10% increase in teachers' wages, relative to their local labour market, was found to improve pupil performance by 0.02sd, but are significantly smaller than those found by Dolton and Marcenaro-Gutierrez (2011), where a 10% increase in teachers' relative wages improves pupil performance by between 0.1sd and 0.2sd. However, this is what we'd expect as Dolton and Marcenaro-Gutierrez (2011) are unable to distinguish between selection effects - countries that pay teachers' higher salaries attract more productive teachers - and efficiency wage effects.

4.3 Labour Market Conditions and Relative wages

Accounting for differences in job security, and the cost of unemployment, using our constructed labour market returns to teaching we find that the coefficients are marginally stronger. Column 6 shows that a 10% increase in the matched labour market returns to teaching causes a 0.03sd increase in Science and 0.024sd in Math (column 13), all statistically significant at the 1% level, while the more general graduates labour market returns (column 7 and 14) show that the effect is 0.026sd and 0.02sd respectively.

The TIMSS assessments are taken between April and June in England. Therefore, our estimates reflect the impact that a more motivated teacher has after 0.8 to 0.9 of an academic year. Therefore, when evaluating the merit of a salary intervention, to improve teacher retention and recruitment, policymakers should also consider the impact on teacher motivation. For example, we estimate that the increase in teachers' pay scales, for the 2020-21 academic year, of 5.5% (the first stage of increasing teachers starting salaries by 24% and more experienced teachers' salaries by 8%) would improve student test scores by roughly 0.016sd in Science and 0.013sd in Math in the first academic year alone, ceteris paribus.

Second our results indicate that, even in the absence of a policy intervention, the fluctuations in teachers' relative wages over the business cycle will impact pupils' test scores. Specifically, during periods of economic downturn (prosperity), pupils will benefit (suffer) from having a more (less) motivated teacher. For example, if the graduate unemployment rate increases by 4% and teachers' salaries rose by 4%, compared to non-teachers, we would expect pupil outcomes to improve by a magnitude quite close to the effect of a 10% increase in teachers' salaries.

4.4 Teacher Pay and Pupil Happiness

A change in teacher effort could also affect their pupils' enjoyment of learning. In the TIMSS students survey students are asked about their attitudes towards learning Mathematics and Science. In response to the question 'I enjoy learning' they can respond Agree a lot, Agree a little, Disagree a little or Disagree a lot. Using this data, we create a dummy that indicates if a young person enjoys leaning the subject (Agree a little or Agree a lot) or not (Disagree a little or Disagree a lot). We find 74% of Grade 4 pupils enjoy learnings science and 80% enjoy learning maths. As the results in table 8 shows, teachers' relative wages also have an effect on their pupil's enjoyment of learning. The main effect is on Science enjoyment (column 1-4) where a 10% increase in the matched labour market returns to teaching increases Science enjoyment by 1.8%, statistically significant at the 1% level.

In the student survey, enjoyment of learning is reported in an ordinal form where 1 indicates that pupils enjoy learning this subject the most and 4 indicates pupils who enjoy learning this subject the least. If we use this variable and regress it on the same covariates using an ordinal probit we find that, in line with our previous results, a 10% increase in teachers' relative wages has a positive effect on Grade 4 pupils enjoying learning Science a lot (1.75%) and has a negative effect on the probability that a Grade 4 pupil does not enjoying learning a lot (-0.85%), all statistically significant at the 10% level (see figure 5).

As the correlation between learning enjoyment and test scores is relatively weak – 0.015 in Science and 0.04 in Math for primary school pupils – it is unlikely that the effect of relative wages on pupil performance is been driven by changes in pupil happiness, and vice versa. A growing body of literature both in England, and abroad, finds that pupils' enjoyment of learning and well-being at school, while unrelated to test score performance, are strong predictors of future labour market success (Gibbons and Silva 2011, Jackson 2012).Therefore, our estimates suggest that relative wages have a causal effect on two distinct outcomes: pupil happiness and pupil performance.

5 Robustness Checks

5.1 Inexperienced Teachers

Assuming that new teachers have strong teaching preferences – they will exert high effort regardless of the outside option – we can test our results to check if they are being driven by teacher effort. We do this by running our OLS model again but excluding new teachers – those whose effort is unlikely to be responsive to variation in the relative wage. We define new teachers as those who have two years of experience or fewer (Table 9).

Using this smaller sample if our coefficients are larger it would suggest that our results are driven by teachers but if they are smaller, or unchanged, it would suggest that our results are driven through some other channel. Consistent with our predictions, restricting our analysis to those teachers whose effort we would expect to be responsive to changes in relative wages increases our effect sizes by 1% of a standard deviation in both Math and Science performance. For example, column 6 in table 9 shows that our effect on Science performance increases when we remove the least experienced teachers (the effect of a 10% increase in wages increases from 0.029sd to 0 0.038sd).

5.2 Academic Attainment Fixed Effects

Ideally we would include region fixed effects in our main model to account for the fact that there are significant regional differences in England that might bias our results. For many countries in TIMSS, such as Australia, Germany and Northern Ireland you could easily do this using the School Strata as the stratification is by region. In England, stratification is done on two levels. The first is by whether the School is just a Primary School or a combined Primary and Secondary school and the second is by the school's prior level of academic attainment. Using the first level we include a dummy for if the school is a Primary School or a combined school. This has no impact on our main results – although Grade 4 pupils in a combined school tend to score .20sd lower in Science and 0.18sd lower in Math.

Apart from 2015, each wave of TIMSS in England is stratified by six levels of the schools' prior level of academic attainment. The prior levels of academic attainment are calculated using key stage 2 results (primary school) and key stage 3 (secondary school). Table 10 shows that pupils in better schools typically achieve higher scores in both Mathematics and Science.[16] For example, students in the best schools typically outperform students from the lowest achieving schools by around one quarter of a standard deviation. Adding academic attainment fixed effects to our model to exploit within year, within similarly achieving schools, variation Table 11 shows that not only do our main results persist, in this more conservative specification, but the effect sizes get marginally larger. Column 3 shows that the effect on Grade 4 Science of a 10% increase in the labour market returns to teaching increases from 0.0296sd to 0.0362sd.

6. Conclusion

Using a novel estimation strategy this paper shows that when we account for selection bias and relative job security there is no strong evidence that teachers could leave teaching for a higher paying occupation. However, we do find that the growth in male teachers' wages tends to be flatter than what they would expect in their outside option. As a consequence, when we take into account the differences in earnings growth there is a high probability (>50%) that a male teacher could maximise their lifetime earnings by leaving the occupation. This is despite the fact that their initial wages are fairly similar. Looking at the earnings of teachers who quit we find no evidence that they tend to leave teaching for higher paying occupations. This is also true for teachers with a degree in a STEM subject who have fairly strong labour market opportunities. This suggests that either teaching is a strong negative signal on the labour market, teachers are misinformed about their outside option or individuals who leave the occupation are not motivated by pecuniary factors.

Using our wage estimates we find that teachers' wages, consistent with an efficiency wage model, improve pupils' test scores and well-being, measured by enjoyment of learning. To put the size of our effect on pupil performance into a policy perspective the magnitude of a

[16] These categories were based on the schools key Stage 2 (KS2) and key stage 3(KS3) results. These are formal assessments that examine young people on the material that they have learnt in years 3 to 6 (ages 6 to 11 (This is KS2)) and year 7 to 9 (ages 11 to 14(KS3)).

10% increases in teachers' relative wages has roughly the same effect that Krueger (1999) found for a 1 pupil reduction in class size in Project STAR and Lavy (2015) found for a one hour increase in weekly instructional time using PISA.

These results indicate that current students will benefit from raising teachers' salaries. Specifically, over an academic year more motivated teachers will improve their students' academic attainment and enjoyment of learning. However, this does not mean that an unconditional salary increase is a cost-effective policy instrument to improve pupil performance since it is extremely expensive. A 10% increase in teachers' relative wages is likely to cost an additional £1.3bn per year in primary schools alone.[17] To put the magnitude of the cost into perspective to achieve the same improvement in pupil performance by reducing class sizes in primary schools would cost £232m.[18] A more efficient mechanism to improve pupil performance could be a conditional wage increase. Atkinson et al., (2009) shows that the effect of performance related pay on pupil performance is noticeably stronger than our estimates and is considerably cheaper to implement.

These results suggests that more experienced teachers are more responsive to wage differentials than less experienced teachers. As the government is committed to increasing less experienced teachers' salaries (roughly 24% by 2022) by significantly more than their more experienced colleagues (8%) this might adversely affect teacher effort. Investigating if teachers' wages, relative to other teachers, influences pupil performance and the potential adverse effects of flattering teachers' pay schedule seems like a promising topic for future research.

This paper provides some evidence that teachers' relative wages also affects pupils' well-being. As well-being plays an important role in a wide range of pupil outcomes failing to consider the wider effects of a policy mechanism might lead to a misallocation of resources (Lévy-Garboua et al., 2006). Therefore, investigating the impact of policy mechanisms on a wider range of outcomes and potential dynamic complementarities seems like an important area of future research.

[17] Using the 2018 SWC 172,055 primary school teaches' (mean salary £38,862) and 83,051 primary academy school teachers (mean salary £37,235). Assumed non-teachers' salaries will grow at 3%.

[18] Reducing primary school class sizes from 27 to 26 would require roughly 9,800 additional teachers. Assuming that we can hire this number of teachers at the lowest point of the pay band (£23,720) and there are not additional costs (such as building additional classrooms or hiring additional support staff).

Table 1 Teacher and non-teacher annual wages, in pounds, adjusted to 2019 prices using the Consumer Price Index, between 1993 and 2019 using the Labor Force Survey.

	(a)	(b)	(c)	(d)
Year	Teacher Wage	Non-teachers Wage		
		Graduates	Matched	Former Teachers
1993	44000	51400	50500	40500
1994	44100	49300	47100	42000
1995	42400	48900	43400	39800
1996	44700	51400	49500	40800
1997	43300	50500	48800	40400
1998	43000	50900	45200	41700
1999	44300	52100	48500	42000
2000	44300	52800	44300	44800
2001	46500	54400	49000	43600
2002	46800	54600	47400	45100
2003	47600	54800	46100	48800
2004	48200	54300	47200	44300
2005	48500	53900	49000	45200
2006	48400	54200	49100	46100
2007	47400	52900	47000	43700
2008	46700	54700	45000	42800
2009	49100	54000	46800	45700
2010	47900	52000	43500	42000
2011	43700	49700	44500	49000
2012	42500	47900	43200	46500
2013	41400	46700	46600	43000
2014	40600	46800	44000	43200
2015	40400	46700	43100	42800
2016	39600	46900	44100	40200
2017	38700	45500	41400	40000
2018	38300	44100	39900	37200
2019	38100	43800	37500	41200

Note: Wages are all rounded to the nearest hundred. Graduates' wages are the average nominal earnings of all non-teaching graduates. Matched Wages are teachers outside option estimated using nearest neighbour propensity score matching by comparing the earnings of teachers to look most similar to graduates based on observable characteristics. Former teachers' wages are the average nominal earnings of all former teachers who remain employed full time.

Table 2 Ratio of teacher and non-teaching wages using our matching strategy and our normal strategy by University Subject Field and years using the LFS.

Years[19]	Strategy	Comparing current teachers to Graduates		Comparing current teachers to qualified teachers who are not teaching	
		(a) STEM	(b) Non-STEM	(c) STEM	(d) Non-STEM
1993-1996	Matching	0.875	0.906	NA	NA
	Normal	0.880	0.872	1.007	1.075
1997-2000	Matching	0.866	0.938	NA	NA
	Normal	0.831	0.869	1.031	1.052
2001-2004	Matching	0.908	1.043	NA	NA
	Normal	0.862	0.887	0.956	1.036
2005-2008	Matching	0.917	1.061	NA	NA
	Normal	0.867	0.911	0.970	1.070
2009-2012	Matching	0.948	1.033	NA	NA
	Normal	0.905	0.922	0.900	0.982
2013-2016	Matching	0.937	0.932	NA	NA
	Normal	0.883	0.869	0.999	0.903

Columns a-b estimate teachers outside option using non-teaching graduates while columns c-d use qualified teachers who are no longer teaching. Columns a and c estimate the outside option for teachers with a degree in a STEM subject and columns b and d estimate it for teachers without a degree in a STEM subject. In columns c and d we are unable to estimate teachers' outside option using propensity score matching using former teachers as our comparison group due to the modest sample size.

[19] To get a sample size large enough to estimate teachers' and non-teachers' wages by degree subject I had to merge 4 years of LFS data together.

Table 3 Difference in student performance, enjoyment of learning and age and sex of students dropped from our sample using 5 waves of TIMSS.

	(1) Grade 4 Keep	(2) Grade 4 Dropped	(3) Grade 8 Keep	(4) Grade 8 Dropped	(5) Grade 8 Keep	(6) Grade 8 Dropped
			Science		Math	
Math Score	529.12	513.05***			513.89	490.94***
	(90.69)	(94.92)			(81.52)	(88.22)
Science Score	534.54	523.00***	551.41	534.18***		
	(82.59)	(90.10)	(84.33)	(85.13)		
Enjoy Math (Dummy)	0.81	0.64***			0.64	0.66***
	(0.396)	(0.479)			(0.481)	(0.473)
Enjoy Science (Dummy)	0.74	0.63***	0.74	0.71***		
	(0.440)	(0.482)	(0.439)	(0.454)		
Student Age	10.04	10.01***	14.15	14.04***	14.15	13.84***
	(0.469)	(0.518)	(0.381)	(0.505)	(0.389)	(0.595)
Student Male	0.50	0.50	0.51	0.49***	0.50	0.51
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
N	25346	3245	17302	9514	15177	4225

Note. Math and Science scores are at the standardized at the TIMSS level with an international mean of 500 and standard deviation of 100. Enjoy Math and Enjoy Science are (dummies where 1 indicates that they enjoy learning or not). Mean coefficients; sd in parentheses, standard errors at the usual levels and indicate statistical significant from the corresponding 'keep' column. For example, stars in column 2 indicate that the mean in column two is statistically different from the mean in column 1.: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0$

Table 4 TIMSS Teachers Wage Descriptive Statistics

	Grade 4		Grade 8			
			Science		Math	
	mean	sd	mean	sd	mean	sd
Ln Teacher Wage	6.316	.249	6.42	.224	6.43	.229
Ln Non Teacher Wage Matched	6.283	.298	6.40	.283	6.42	.283
Ln Non Teacher Wage Graduate	6.384	.298	6.52	.282	6.53	.283
Difference In Wage Matched	.032	.123	.009	.123	.004	.122
Difference in Wage Graduate	-.068	.101	-.104	.105	-.104	.104
Teacher Unemployment Rate	1.710	.438	1.73	.469	1.76	.499
Graduate Unemployment Rate	3.123	1.400	2.94	1.38	2.93	1.40
Labour Market Differences Match	.061	.126	.034	.126	.029	.126
Labour Market Differences Graduate	-.037	.111	-.076	.114	-.077	.113
N	25,346		17,302		15,177	

Note. The estimates for teachers' and non-teachers' wages come from 1993-2019 LFS with all wages adjusted to 2019 prices. Non-teacher Wage graduates is the average non-teaching graduates wage while non-teacher wage matched is non-teaching graduates' wage matched to teachers using nearest neighbour propensity score matching. The difference in wages is Log(Teacher Wage) – Log (Non-Teacher Wage) while the labour market differences is the same but they define *Log(Teacher Wage) as* $P(Emp|Teacher, Age, Sex) Log\ w_{Teacher} + \left(1 - P(Emp|Teacher, Age, Sex)\right) Log\ w_{JSA}$ *and Log (Non-Teacher Wage) as* $P(Emp|Non-Teacher, Age, Sex) Log\ w_{Non-Teacher} + (1 - P(Emp|Non-Teacher, Age, Sex)) Log\ w_{JSA}$.

Table 5 TIMSS students descriptive statistics.						
	Grade 4		Grade 8			
			Science		Math	
	mean	sd	mean	sd	mean	sd
Student Age	10.040	.469	14.148	.381	14.147	.3893
Student Male	.497	.500	.507	.499	.499	.500
Books at home:						
0-10	.097	.295	.119	.324	.135	.342
11-25	.189	.391	.181	.385	.197	.398
26-100	.320	.466	.286	.452	.286	.452
101-200	.198	.399	.191	.393	.184	.388
200+	.195	.396	.220	.414	.196	.397
Home Computer	.870	.335	.942	.233	.947	.225
Speak English in Home	.781	.414	.868	.337	.873	.332
Class Size Above Median						
Math	.540	.498			.542	.498
Science	.549	.498	.534	.498		
Teacher experience Years:						
1	.092	.289	.078	.269	.078	.268
2	.100	.300	.071	.258	.071	.258
3	.078	.269	.065	.247	.076	.265
4	.076	.252	.063	.244	.053	.224
5	.087	.282	.044	.206	.050	.218
6+	.565	.495	.675	. 468	.670	.470
Techer Age:						
Under 25	.078	.268	.049	.218	.061	.240
25-29	.087	.282	.191	.393	.167	.373
30-39	.086	.280	.300	.458	.270	.444
40-49	.197	.398	.237	.425	.278	.448
50-59	.276	.447	.199	.399	.196	.397
60+	.154	.362	.021	.143	.025	.158
Teacher Male	.261	.439	.495	.499	.497	.500
n	25,346		17,302		15,177	

This table shows the descriptive statistics of the students in our TIMSS sample. For example the final row should the proportion of Grade 4 students who have male teachers (column 1, 26.1%).

Table 6 OLS regression of Grade 4 and 8 pupil performance on observable characteristics in TIMSS.

	1	2	3	4
	Grade 4		Grade 8	
	Math Score	Science Score	Math Score	Science Score
Class Size Above Median	0.123***	0.101**	0.557***	0.271***
	(0.0445)	(0.0401)	(0.0555)	(0.0519)
Relative Student Age	0.0350***	0.0364***	0.0126***	0.0126***
	(0.00230)	(0.00237)	(0.00206)	(0.00187)
Student Male	0.0876***	0.0489***	0.0924***	0.148***
	(0.0155)	(0.0151)	(0.0222)	(0.0169)
books at home 0-10 (Omitted)				
books at home 11-25	0.381***	0.448***	0.361***	0.440***
	(0.0282)	(0.0268)	(0.0266)	(0.0257)
books at home 26-100	0.748***	0.805***	0.705***	0.809***
	(0.0281)	(0.0275)	(0.0320)	(0.0274)
books at home 101-200	0.994***	1.083***	0.961***	1.176***
	(0.0314)	(0.0313)	(0.0382)	(0.0322)
books at home 200+	1.059***	1.246***	1.232***	1.493***
	(0.0342)	(0.0333)	(0.0431)	(0.0329)
Computer in Home	-0.0388+	-0.000321	0.0540	-0.0134
	(0.0263)	(0.0254)	(0.0377)	(0.0314)
Speak English in Home	0.0103	0.150***	-0.106***	0.0371+
	(0.0243)	(0.0238)	(0.0321)	(0.0249)
Teacher Male	0.0180	0.0236	0.00976	0.0514*
	(0.0315)	(0.0291)	(0.0440)	(0.0290)
Teacher Experience 1 Year	-0.0910*	-0.0888**	-0.0398	0.0317
	(0.0465)	(0.0427)	(0.0789)	(0.0657)
Teacher Experience 2 Years	0.0256	0.0148	0.0207	-0.0549
	(0.0591)	(0.0526)	(0.0984)	(0.0665)
Teacher Experience 3 Years	0.0146	0.00716	-0.191**	0.0159
	(0.0542)	(0.0493)	(0.0902)	(0.0618)
Teacher Experience 4 Years	-0.0834+	-0.0709+	0.163	-0.00617
	(0.0511)	(0.0489)	(0.115)	(0.0620)
Teacher Experience 5 Years	-0.0162	0.00698	-0.0427	-0.0878
	(0.0593)	(0.0563)	(0.117)	(0.0702)
Teacher Experience 6+ years (Omitted)				

[continues on next page]

Teacher age Under 25	-0.0752	-0.0689	-0.00211	-0.215***
	(0.0560)	(0.0563)	(0.111)	(0.0769)
Teacher age 25-29	0.0263	0.00820	0.0765	-0.0315
	(0.0470)	(0.0429)	(0.0746)	(0.0508)
Teacher age 30-39	-0.0157	-0.0285	0.0254	-0.00447
	(0.0347)	(0.0338)	(0.0628)	(0.0382)
Teacher Age 40-49 (Omitted)				
Teacher Age 50-59	0.0984**	0.0992**	-0.110+	0.0183
	(0.0414)	(0.0388)	(0.0681)	(0.0449)
Teacher Age 50+	0.412***	0.230	-0.0295	0.114
	(0.154)	(0.168)	(0.164)	(0.107)
Constant	-0.589***	-0.922***	-0.875***	-1.087***
	(0.0658)	(0.0598)	(0.0880)	(0.0749)
N	25346	25366	15177	17302

Our Dependent variables are standardized to have a mean of 0 and a standard deviation of 1. The regressions include year dummies but are not reported. Standard errors are clustered at the class level and the starts indicate statistical significance at the following levels: +p<0.15, *p<0.10, **p<0.05, ***p<0.01.

Table 7 OLS regression of grade 4 standardized science and math scores in TIMSS on teachers wages

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
				Science							Math			
Log Teacher Wages	0.240 (0.244)	0.417* (0.239)	0.447+ (0.273)					-0.0801 (0.232)	0.0843 (0.238)	0.129 (0.266)				
Log Non-Teacher Wages (Match)		-0.259*** (0.0824)							-0.241*** (0.0771)					
Log Non-Teacher Wages (Normal)			-0.208+ (0.139)							-0.209+ (0.130)				
Wage Difference (Match)				0.265*** (0.0799)							0.235*** (0.0766)			
Wage Difference (Normal)					0.208+ (0.139)							0.209+ (0.130)		
Labor Market Returns to Teaching (Match)						0.296*** (0.0840)							0.242*** (0.0796)	
Labor Market Returns to Teaching (Norm)							0.259* (0.142)							0.197+ (0.132)
Constant	-0.895*** (0.0304)	0.810+ (0.545)	0.489 (0.925)	-0.897*** (0.0304)	-0.896*** (0.0304)	-0.896*** (0.0304)	-0.896*** (0.0304)	-0.564*** (0.0309)	1.025** (0.509)	0.827 (0.864)	-0.565*** (0.0309)	-0.564*** (0.0309)	-0.565*** (0.0309)	-0.564*** (0.0309)
N	25346	25346	25346	25346	25346	25346	25346	25346	25346	25346	25346	25346	25346	25346

Our Dependent variables are standardized to have a mean of 0 and a standard deviation of 1. Regression includes all of our controls, these are: Class Size, Class Size Missing Dummy, Student Age above median in months, student sex, books at home, computer in home, speak English at home, teacher sex, teacher experience, teacher age and year fixed effects. Standard Errors in parentheses clustered at the class room level. Signs indicate significance at the following level +p<0.15,* p<0.10,**p<0.05,***p<0.01. Note: standard errors obtained from bootstrap (500)

Table 8 OLS regression of grade 4 students enjoyment of learnings in TIMSS on teachers wages

	1	2	3	4	5	6	7	8
	Grade 4							
	Science				Math			
Wage Difference (Match)	0.169*** (0.0367)				-0.0122 (0.0325)			
Wage Difference (Normal)		0.197*** (0.0612)				0.0285 (0.0553)		
Labor Market Returns to Teaching (Match)			0.182*** (0.0379)				-0.0109 (0.0359)	
Labor Market Returns to Teaching (Norm)				0.215*** (0.0629)				0.0334 (0.0563)
Constant	0.787*** (0.0151)	0.787*** (0.0152)	0.787*** (0.0151)	0.787*** (0.0152)	0.849*** (0.0128)	0.849*** (0.0128)	0.849*** (0.0127)	0.849*** (0.0127)
DP mean	.757	.757	.757	.757	.820	.820	.820	.820
(SD)	(.428)	(.428)	(.428)	(.428)	(.383)	(.383)	(.383)	(.383)
N	24659	24659	24659	24659	24872	24872	24872	24872

Regression includes all of our controls, these are: Class Size, Student Age above median in months, student sex, books at home, computer in home, speak English at home, teacher sex, teacher experience, teacher age and year fixed effects. Standard Errors in parentheses clustered at the classroom level. Signs indicate significance at the following levels +p<0.15,* p<0.10,**p<0.05,***p<0.01. Note: standard errors obtained from bootstrap (500) and our sample size is marginally smaller because 474 pupils (1.8%) did not complete this question.

Table 9 OLS regression of grade 4 Science and Math scores excluding teachers with two or less years experiences in TIMSS on teachers wages

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
				Science							Math			
Log Teacher Wages	0.279	0.536**	0.526*					-0.223	-0.00275	-0.0252				
	(0.267)	(0.272)	(0.310)					(0.259)	(0.264)	(0.291)				
Log Non-Teacher Wages (Match)		-0.344***							-0.297***					
		(0.0994)							(0.0930)					
Log Non-Teacher Wages (Normal)			-0.270*							-0.216+				
			(0.159)							(0.148)				
Wage Difference (Match)				0.351***							0.287***			
				(0.0985)							(0.0921)			
Wage Difference (Normal)					0.277*							0.209		
					(0.159)							(0.148)		
Labor Market Returns to Teaching (Match)						0.386***							0.299***	
						(0.104)							(0.0969)	
Labor Market Returns to Teaching (Norm)							0.332**							0.206
							(0.165)							(0.151)
Constant	-0.903***	1.367**	0.896	-0.906***	-0.905***	-0.906***	-0.904***	-0.570***	1.389**	0.868	-0.572***	-0.571***	-0.572***	-0.570***
	(0.0344)	(0.657)	(1.064)	(0.0344)	(0.0344)	(0.0344)	(0.0344)	(0.0353)	(0.613)	(0.987)	(0.0353)	(0.0353)	(0.0353)	(0.0353)
N	20462	20462	20462	20462	20462	20462	20462	20462	20462	20462	20462	20462	20462	20462

Our Dependent variables are standardized to have a mean of 0 and a standard deviation of 1. Regression includes all of our controls, these are: Class Size, Class Size Missing Dummy, Student Age above median in months, student sex, books at home, computer in home, speak English at home, teacher sex, teacher experience, teacher age and year fixed effects. Signs indicate significance at the following levels. Standard Errors in parentheses. +p<0.15,* p<0.10,**p<0.05,***p<0.01. Note: standard errors obtained from bootstrap (500)

Table 10 OLS regression of primary school pupils' Math and Science scores on schools academic attainment levels in TIMSS

	1 Science Score	2 Math Score
Attainment Level 1 (Omitted)		
Attainment Level 2	0.0757+	0.0610
	(0.0488)	(0.0495)
Attainment Level 3	0.0873*	0.0541
	(0.0484)	(0.0470)
Attainment Level 4	0.156***	0.175***
	(0.0540)	(0.0524)
Attainment Level 5	0.166***	0.133***
	(0.0516)	(0.0468)
Attainment Level 6	0.236***	0.230***
	(0.0548)	(0.0559)
_cons	-1.244***	-0.870***
	(0.0752)	(0.0756)
N	17951	17951

Our Dependent variables are standardized to have a mean of 0 and a standard deviation of 1. Regression includes all of our controls, these are: Class Size, Student Age above median in months, student sex, books at home, computer in home, speak English at home, teacher sex, teacher experience, teacher age and year fixed effects. Standard Errors in parentheses clustered at the class level and significant is displayed at the usual levels. +p<0.15,* p<0.10,**p<0.05,***p<0.01. Note these attainment categories are not ordered in TIMSS, I ordered and named them based on the pupils' science scores where the category with the lowest scores is 1 and highest is 6. Also note that the sample sizes are slightly smaller as this table excludes the 2015 survey as the prior attainment data is unavailable.

Table 11 OLS regression of Grade 4 Math and Science Scores using schools prior attainment fixed effects in TIMSS on teachers wages

	1	2	3	4	5	6	7	8
	Science				Math			
Wage Difference (Match)	0.331*** (0.125)				0.255** (0.107)			
Wage Difference (Normal)		0.152 (0.176)				0.0886 (0.167)		
Labor Market Returns to Teaching (Match)			0.362*** (0.124)				0.255** (0.106)	
Labor Market Returns to Teaching (Norm)				0.255+ (0.176)				0.113 (0.167)
constant	-1.198*** (0.0449)	-1.176*** (0.0448)	-1.199*** (0.0449)	-1.180*** (0.0448)	-0.831*** (0.0416)	-0.813*** (0.0415)	-0.830*** (0.0416)	-0.814*** (0.0415)
N	17931	17931	17931	17931	17931	17931	17931	17931

Our Dependent variables are standardized to have a mean of 0 and a standard deviation of 1. Regression includes all of our controls, these are: Class Size, Student Age above median in months, student sex, books at home, computer in home, speak English at home, teacher sex, teacher experience, teacher age and year fixed effects. Standard Errors in parentheses level and significant is displayed at the usual levels. +p<0.15,* p<0.10,**p<0.05,***p<0.01. . Note: standard errors obtained from bootstrap (500). Our sample size is significantly lower using School prior attainment FE's because the prior attainment data is unavailable in the 2015 survey.

3.7 Figures

Figure 1. Average teachers' Pay between 1993 and 2019 as a ratio of graduates pay

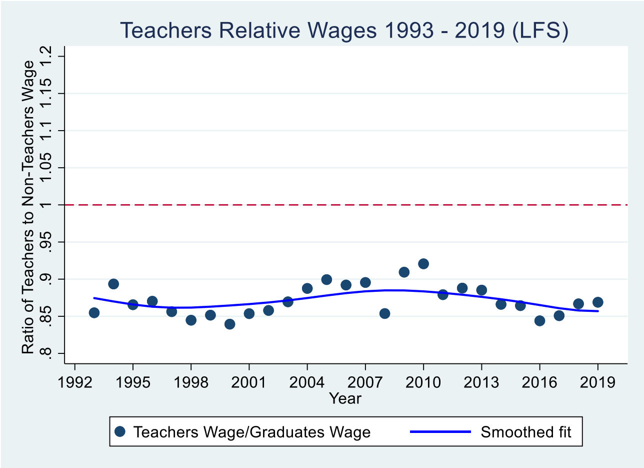


Figure 2 Average teachers' pay between 1993- 2019 as a ratio of graduates pay by age. The LHS is younger teachers and graduates (under 30 and 30-39) and the RHS is older teachers (40-49 and 50-59).

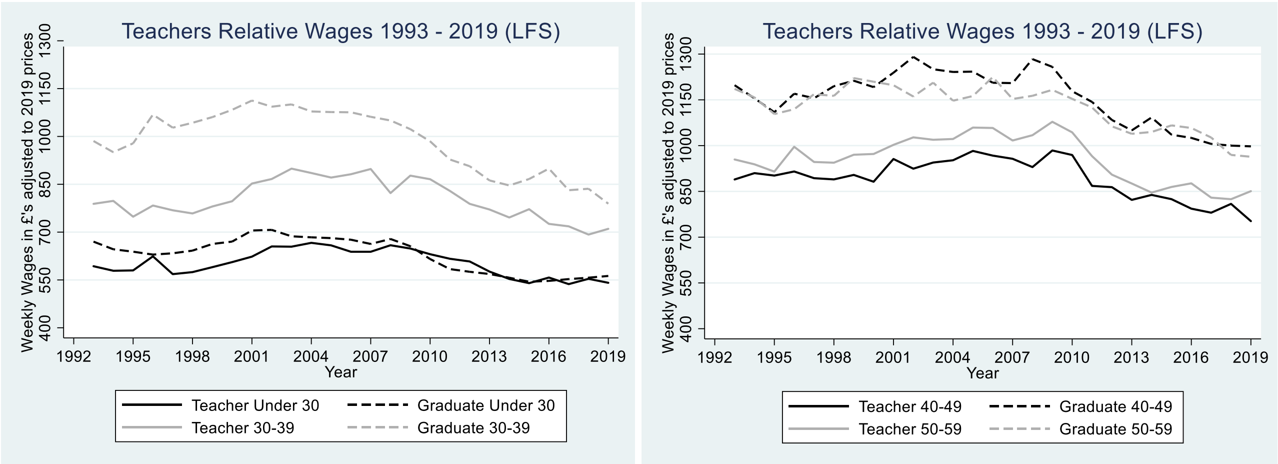
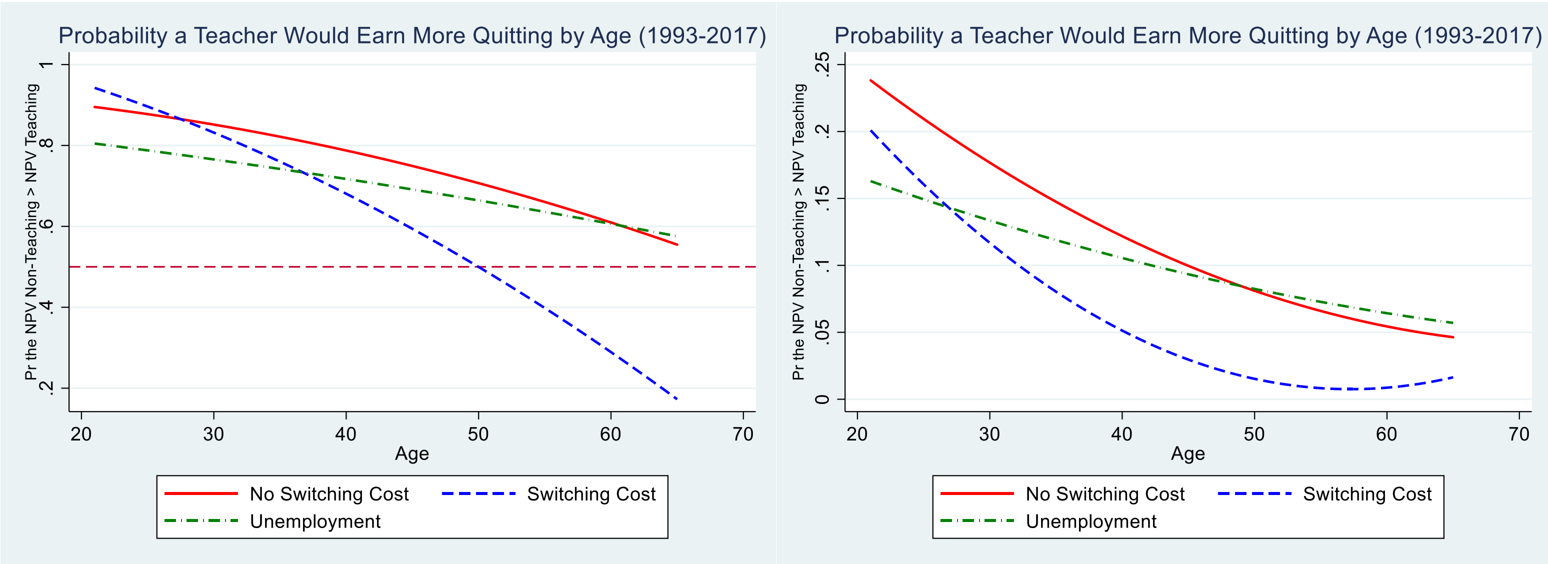
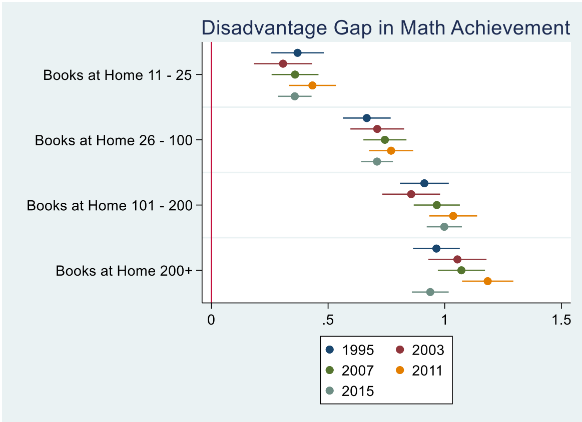


Figure 3 shows the probability that a teacher quitting would maximise their lifetime earnings by age and sex (Male LHS and Female RHS) using the high discounting parameter (25%). The red solid line assumes that markets perfectly clear (i.e. an individual is employed as a teacher or non-teacher with probability 1) and no switching cost. The Blue dashed line assumes that markets perfectly clear but there is a switching cost of 10% (i.e. when teaching sort out of teaching they face an immediate pay penalty). Finally the Green dash dot line is the same as the solid red li ne but relaxes the assumption about perfect market clearance using the teachers and non-teachers actual unemployment rates from the Labour Force Survey. See the supplementary material for more information.



These figures show clear differences in quitting intentions by male (LHS) and female (RHS) teachers. Even with a high switching cost the probability that a male teacher could maximise their lifetime earnings by leaving the occupation exceeds 50% for the majority of their career wh ile for female teachers the probability is significantly less likely.

Figure 4 shows the change in the difference in achievement by our SES proxy "Books at Home" in Grade 4 Math (LHS) and Science (RHS) achievement in a multivariate regression including all our usual controls. The Omitted variable is 0 – 10 Books at Home.



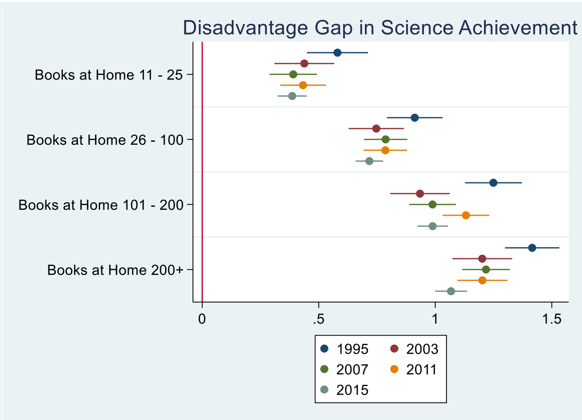
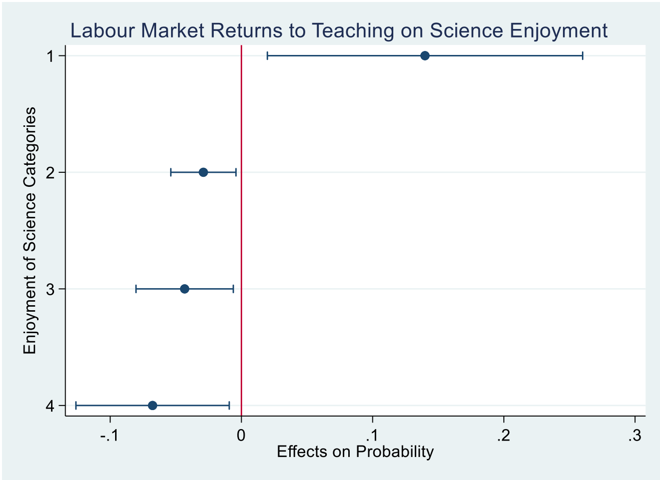


Figure 5 shows the marginal effect of a 1% increase in the labour market returns to teaching on grade 4 science enjoyment whe re category 1 is enjoy learning the most and category 4 is enjoy learning the least. The confidence intervals are at the 90% level.



References

AKERLOF, GEORGE A (1982): "Labor Contracts as Partial Gift Exchange," *The Quarterly Journal of Economics*, 97, 543-569.

ATKINSON, ADELE, SIMON BURGESS, BRONWYN CROXSON, PAUL GREGG, CAROL PROPPER, HELEN SLATER, and DEBORAH WILSON (2009): "Evaluating the Impact of Performance-Related Pay for Teachers in England," *Labour Economics*, 16, 251-261.

BEDARD, KELLY, and ELIZABETH DHUEY (2006): "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects," *The Quarterly Journal of Economics*, 121, 1437-1472.

BRITTON, JACK, and CAROL PROPPER (2016): "Teacher Pay and School Productivity: Exploiting Wage Regulation," *Journal of Public Economics*, 133, 75-89.

BRYSON, ALEX, and KERRY L PAPPS (2016): "Spillovers and Team Incentives," Working Paper, University College, London, and University of Bath.

CALIENDO, MARCO, and SABINE KOPEINIG (2008): "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of economic surveys*, 22, 31-72.

CHETTY, RAJ, JOHN N FRIEDMAN, and JONAH E ROCKOFF (2011): "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood," National Bureau of Economic Research.

CHEVALIER, ARNAUD, and PETER J DOLTON (2004): "The Labour Market for Teachers," Working Paper Series.

CHEVALIER, ARNAUD, PETER DOLTON, and STEVEN MCINTOSH (2007): "Recruiting and Retaining Teachers in the Uk: An Analysis of Graduate Occupation Choice from the 1960s to the 1990s," *Economica*, 74, 69-96.

CONTINI, DALIT, MARIA LAURA DI TOMMASO, and SILVIA MENDOLIA (2017): "The Gender Gap in Mathematics Achievement: Evidence from Italian Data," *Economics of Education Review*, 58, 32-42.

DAHL, GORDON B, and LANCE LOCHNER (2012): "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit," *American Economic Review*, 102, 1927-56.

DE REE, JOPPE, KARTHIK MURALIDHARAN, MENNO PRADHAN, and HALSEY ROGERS (2015): "Double for Nothing? Experimental Evidence on the Impact of an Unconditional Teacher Salary Increase on Student Performance in Indonesia," National Bureau of Economic Research.

DOLTON, PETER (2004): "Recruiting High Quality Teachers," *CESifo DICE Report*, 2, 11-20.

DOLTON, PETER, and OSCAR D MARCENARO-GUTIERREZ (2011): "If You Pay Peanuts Do You Get Monkeys? A Cross-Country Analysis of Teacher Pay and Pupil Performance," *Economic policy*, 26, 5-55.

DOLTON, PETER, ANDREW TREMAYNE, and TSUNG-PING CHUNG (2003): *The Economic Cycle and Teacher Supply*. OECD Paris.

DUNCAN, GREG J, and RICHARD J MURNANE (2011): *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*. Russell Sage Foundation.

FEHR, ERNST, LORENZ GOETTE, and CHRISTIAN ZEHNDER (2009): "A Behavioral Account of the Labor Market: The Role of Fairness Concerns," *Annu. Rev. Econ.*, 1, 355-384.

FRIEDMAN, MILTON, and SIMON KUZNETS (1945): *Income from Independent Professional*. New York: National Bureau of Economic Research.

FULLARD, JOSHUA (2018): "“Bad Economy, Good Teachers. The Countercyclicality of Enrolment into Initial Teacher Training Programmes in the Uk”," Institute for Social and Economic Research.

FULLARD, JOSHUA (2020a): "Teacher Supply and Covid-19," https://epi.org.uk/publications-and-research/teacher-supply-and-covid-19/: Education Policy Institute.

FULLARD, JOSHUA (2020b): "Trends in the Diversity of Teachers in England," https://epi.org.uk/publications-and-research/diversity-of-teachers/: Education Policy Institute.

FULLARD, JOSHUA (2021a): "The Pandemic and Teacher Attrition: An Exodus Waiting to Happen?," https://epi.org.uk/publications-and-research/the-pandemic-and-teacher-attrition-an-exodus-waiting-to-happen/: Education Policy Institute.
FULLARD, JOSHUA (2021b): "Teacher Pay Policy: One Size Fits All? ," https://epi.org.uk/publications-and-research/teacher-pay-policy-blog/: Education Policy Institute.
FULLARD, JOSHUA, and JAMES ZUCCOLLO (2021): "Local Pay and Teacher Retention in England."
GIBBONS, STEPHEN, and OLMO SILVA (2011): "School Quality, Child Wellbeing and Parents' Satisfaction," *Economics of Education Review*, 30, 312-331.
GREANY, TOBY, IAIN BARNES, TAREK MOSTAFA, NICOLA PENSIERO, and CHRISTINA SWENSSON (2016): "Trends in Maths and Science Study (Timss): National Report for England."
HANUSHEK, ERIC A (1986): "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of economic literature*, 24, 1141-1177.
HANUSHEK, ERIC A (2004): "Some Simple Analytics of School Quality," National Bureau of Economic Research.
HANUSHEK, ERIC A (2011a): "The Economic Value of Higher Teacher Quality," *Economics of Education Review*, 30, 466-479.
HANUSHEK, ERIC A (2011b): "How Much Is a Good Teacher Worth," *Education next*, 11, 40-45.
HANUSHEK, ERIC A, JOHN F KAIN, and STEVEN G RIVKIN (1999): "Do Higher Salaries Buy Better Teachers?," National bureau of economic research.
HANUSHEK, ERIC A, PAUL E PETERSON, LAURA M TALPEY, and LUDGER WOESSMANN (2019): "The Unwavering Ses Achievement Gap: Trends in Us Student Performance," National Bureau of Economic Research.
HANUSHEK, ERIC A, JENS RUHOSE, and LUDGER WOESSMANN (2015): "Economic Gains for Us States from Educational Reform," National Bureau of Economic Research.
HECKMAN, JAMES J, HIDEHIKO ICHIMURA, and PETRA E TODD (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64, 605-654.
HEINZ, MANUELA (2015): "Why Choose Teaching? An International Review of Empirical Studies Exploring Student Teachers' Career Motivations and Levels of Commitment to Teaching," *Educational Research and Evaluation*, 21, 258-297.
HEISSEL, JENNIFER, CLAUDIA PERSICO, and DAVID SIMON (2019): "Does Pollution Drive Achievement? The Effect of Traffic Pollution on Academic Performance," National Bureau of Economic Research.
HERMANN, ZOLTÁN, and ALFA DIALLO (2017): "Does Teacher Gender Matter in Europe? Evidence from Timss Data," Budapest Working Papers on the Labour Market.
HIJZEN, ALEXANDER, RICHARD UPWARD, and PETER W WRIGHT (2010): "The Income Losses of Displaced Workers," *Journal of Human Resources*, 45, 243-269.
JACKSON, C KIRABO (2012): "Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina," National Bureau of Economic Research.
KINGDON, GEETA GANDHI, and FRANCIS TEAL (2007): "Does Performance Related Pay for Teachers Improve Student Performance? Some Evidence from India," *Economics of Education Review*, 26, 473-486.
KINSLER, JOSH (2016): "Teacher Complementarities in Test Score Production: Evidence from Primary School," *Journal of Labor Economics*, 34, 29-61.
KRUEGER, ALAN B (1999): "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, 114, 497-532.
LAVY, VICTOR (2015): "Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries," *The Economic Journal*, 125, F397-F424.
LECHNER, MICHAEL (2004): "Sequential Matching Estimation of Dynamic Causal Models."

LÉVY-GARBOUA, LOUIS, YOUENN LOHEAC, and BERTRAND FAYOLLE (2006): "Preference Formation, School Dissatisfaction and Risky Behavior of Adolescents," *Journal of Economic Psychology*, 27, 165-183.

LOYALKA, PRASHANT, SEAN SYLVIA, CHENGFANG LIU, JAMES CHU, and YAOJIANG SHI (2019): "Pay by Design: Teacher Performance Pay Design and the Distribution of Student Achievement," *Journal of Labor Economics*, 37, 621-662.

MANSKI, CHARLES F (1987): "Academic Ability, Earnings, and the Decision to Become a Teacher: Evidence from the National Longitudinal Study of the High School Class of 1972," in *Public Sector Payrolls*: University of Chicago Press, 291-316.

MEYER, ANDREW (2013): "Estimating Discount Factors for Public and Private Goods and Testing Competing Discounting Hypotheses," *Journal of Risk and Uncertainty*, 46, 133-173.

MUÑOZ, JUAN SEBASTIÁN (2018): "The Economics Behind the Math Gender Gap: Colombian Evidence on the Role of Sample Selection," *Journal of Development Economics*, 135, 368-391.

NAGLER, MARKUS, MARC PIOPIUNIK, and MARTIN R WEST (2015): "Weak Markets, Strong Teachers: Recession at Career Start and Teacher Effectiveness," National Bureau of Economic Research.

NICKELL, STEPHEN, and GLENDA QUINTINI (2002): "The Consequences of the Decline in Public Sector Pay in Britain: A Little Bit of Evidence," *The Economic Journal*, 112.

PRIYADHARSHINI, ESTHER, and ANNA ROBINSON-PANT (2003): "The Attractions of Teaching: An Investigation into Why People Change Careers to Teach," *Journal of Education for Teaching*, 29, 95-112.

RAFF, DANIEL MG, and LAWRENCE H SUMMERS (1987): "Did Henry Ford Pay Efficiency Wages?," *Journal of Labor Economics*, 5, S57-S86.

RICKMAN, BILL D, and CARL D PARKER (1990): "Alternative Wages and Teacher Mobility: A Human Capital Approach," *Economics of Education Review*, 9, 73-79.

RIVKIN, STEVEN G, ERIC A HANUSHEK, and JOHN F KAIN (2005): "Teachers, Schools, and Academic Achievement," *Econometrica*, 73, 417-458.

ROSENBAUM, PAUL R, and DONALD B RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

SCAFIDI, BENJAMIN, DAVID L SJOQUIST, and TODD R STINEBRICKNER (2006): "Do Teachers Really Leave for Higher Paying Jobs in Alternative Occupations?," *Advances in Economic Analysis & Policy*, 6.

SHAPIRO, CARL, and JOSEPH E STIGLITZ (1984): "Equilibrium Unemployment as a Worker Discipline Device," *The American Economic Review*, 74, 433-444.

SHARIR, SHMUEL, and YORAM WEISS (1974): "The Role of Absolute and Relative Wage Differentials in the Choice of a Work Activity," *Journal of Political Economy*, 82, 1269-1275.

SMITHERS, ALAN, and PAMELA ROBINSON (2003): "Factors Affecting Teachers' Decisions to Leave the Profession."

STINEBRICKNER, TODD R (1998): "An Empirical Investigation of Teacher Attrition," *Economics of Education Review*, 17, 127-136.

STRØM, BJARNE (2004): "Student Achievement and Birthday Effects," *Unpublished manuscript, Norwegian University of Science and Technology*.

SUN, MIN, SUSANNA LOEB, and JASON A GRISSOM (2017): "Building Teacher Teams: Evidence of Positive Spillovers from More Effective Colleagues," *Educational Evaluation and Policy Analysis*, 39, 104-125.

VIGNOLES, ANNA, AUGUSTIN DE COULON, and OSCAR MARCENARO-GUTIERREZ (2011): "The Value of Basic Skills in the British Labour Market," *Oxford Economic Papers*, 63, 27-48.

WEBB, RICHARD, and SOFÍA VALENCIA (2005): "Human Resources in Public Health and Education in Peru," *An Opportunity for a Different*, 697.

WILLIAMSON, ELIZABETH, RUTH MORLEY, ALAN LUCAS, and JAMES CARPENTER (2012): "Propensity Scores: From Naive Enthusiasm to Intuitive Understanding," *Statistical methods in medical research*, 21, 273-293.

WOESSMANN, LUDGER (2011): "Cross-Country Evidence on Teacher Performance Pay," *Economics of Education Review*, 30, 404-418.

WOESSMANN, LUDGER, and MARTIN R WEST (2002): "Class-Size Effects in School Systems around the World: Evidence from between-Grade Variation in Timss," Institute for the Study of Labor (IZA).

ZHANG, JIAN, SONGQING JIN, and WEI SI (2019): "Incentivising Teachers? Evaluating the Incentive Effect of China's Teacher Performance-Based Compensation Reform in Rural China," *Australian Journal of Agricultural and Resource Economics*.

Appendix

Appendix Table 1 shows the differences in observable characteristics between graduates who go into teaching and those who do not for the years 2000 (column a vs b) and 2010 (d vs e) and how using propensity score matching reduces the observable difference between teachers and non-teachers (a vs c and d vs f).

	(a)	(b)	(c)	(d)	(e)	(f)
		Year 2000			Year 2010	
Variable	Teachers	Non-Teachers		Teachers	Non-Teachers	
		Unmatched	Matched		Unmatched	Matched
Male	.417	.65***	.45*	.369	.588***	.381
White	.967	.928***	.968	.946	.869***	.933
Age	41.46	36.85***	42.08*	41.5	39.1***	42.2
Married	.669	.538***	.649	.626	.561***	.601
Region of Domicile:						
London	.117	.221***	.110	.118	.193***	.125
South East	.284	.310**	.295	.300	.281	.287
Degree Subjects:						
Medicine	.017	.082***	.017	.018	.101***	.015
Education	.414	.022***	.412	.437	.024***	.433
Mathematical Sciences	.151	.307***	.147	.137	.270***	.132
Biological Sciences	.062	.064	.067	.071	.079	.072
Social Sciences	.113	.369***	.117	.108	.380***	.107
Humanities	.186	.109***	.182	.168	.085***	.178
Art	.053	.042*	.052	.058	.045**	.059
n	1,573	6,400		1,459	7,409	

The starts indicate statistical significance in the difference in observable characteristics between the non-teachers (columns b, c and e,f) and teachers (columns a and d respectively)to the usual levels * p<0.10,**p<0.05,***p<0.01
The data source is the 2000 and 2010 labour force surveys. The sample is restricted to graduates who work full time and are between the ages of 21 and 65. Teachers (column a and d) are teachers who teach in a primary or secondary school. Non-teachers (column b,c and e,f) are defined are any non-teaching graduate.

Appendix Table 2 The impact of teachers wages on Grade 8 Scores in TIMSS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
				Science							Math			
Log Teacher Wages	0.170 (0.268)	-0.0235 (0.286)	0.454 (0.341)					-0.674** (0.264)	-0.720*** (0.275)	0.390 (0.339)				
Log Non-Teacher Wages (Match)		0.226 (0.225)							0.0611 (0.324)					
Log Non-Teacher Wages (Normal)			-0.250 (0.181)							-0.955* (0.557)				
Wage Difference (Match)				-0.222* (0.113)							-0.0918 (0.118)			
Wage Difference (Normal)					0.236 (0.180)							0.987*** (0.187)		
Labor Market Returns to Teaching (Match)						-0.200* (0.115)							-0.0444 (0.121)	
Labor Market Returns to Teaching (Norm)							0.256+ (0.176)							1.056*** (0.186)
Constant	-1.087*** (0.0455)	-2.584*** (0.752)	0.584 (1.213)	-1.087*** (0.0454)	-1.086*** (0.0455)	-1.087*** (0.0454)	-1.086*** (0.0455)	-0.876*** (0.0415)	-1.280+ (0.789)	5.505*** (1.255)	-0.875*** (0.0415)	-0.878*** (0.0415)	-0.875*** (0.0415)	-0.877*** (0.0415)
N	17302	17302	17302	17302	17302	17302	17302	15177	15177	15177	15177	15177	15177	15177

Our Dependent variables are standardized to have a mean of 0 and a standard deviation of 1. Regression includes all of our controls, these are: class size, student age above median in months, student sex, books at home, computer in home, speak English at home, teacher sex, teacher experience, teacher age and year fixed effects. Standard errors in parentheses clustered at the classroom level while statistical significant is indicated by: +p<0.15,* p<0.10,**p<0.05,***p<0.01. . Note: standard errors obtained from bootstrap (500).

Appendix Table 3 The impact of teachers wages on Grade 8 Scores Excluding teachers with two or less years experiences in TIMSS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
				Science							Math			
Log Teacher Wages	0.218	0.0464	0.553+					-0.630**	-0.871***	0.223				
	(0.302)	(0.316)	(0.365)					(0.306)	(0.316)	(0.351)				
Log Non-Teacher Wages (Match)		0.199							0.287					
		(0.229)							(0.355)					
Log Non-Teacher Wages (Normal)			-0.319							-0.820				
			(0.430)							(0.620)				
Wage Difference (Match)				-0.194+							-0.303**			
				(0.125)							(0.127)			
Wage Difference (Normal)					0.314*							0.830***		
					(0.189)							(0.190)		
Labor Market Returns to Teaching (Match)						-0.177							-0.223*	
						(0.125)							(0.129)	
Labor Market Returns to Teaching (Norm)							0.319*							0.987***
							(0.188)							(0.194)
Constant	-1.063***	-2.379***	1.070	-1.065***	-1.063***	-1.065***	-1.063***	-0.915***	-2.813***	4.565***	-0.914***	-0.915***	-0.914***	-0.914***
	(0.0527)	(0.833)	(1.274)	(0.0527)	(0.0528)	(0.0527)	(0.0528)	(0.0479)	(0.850)	(1.271)	(0.0480)	(0.0479)	(0.0480)	(0.0479)
N	14696	14696	14696	14696	14696	14696	14696	12901	12901	12901	12901	12901	12901	12901

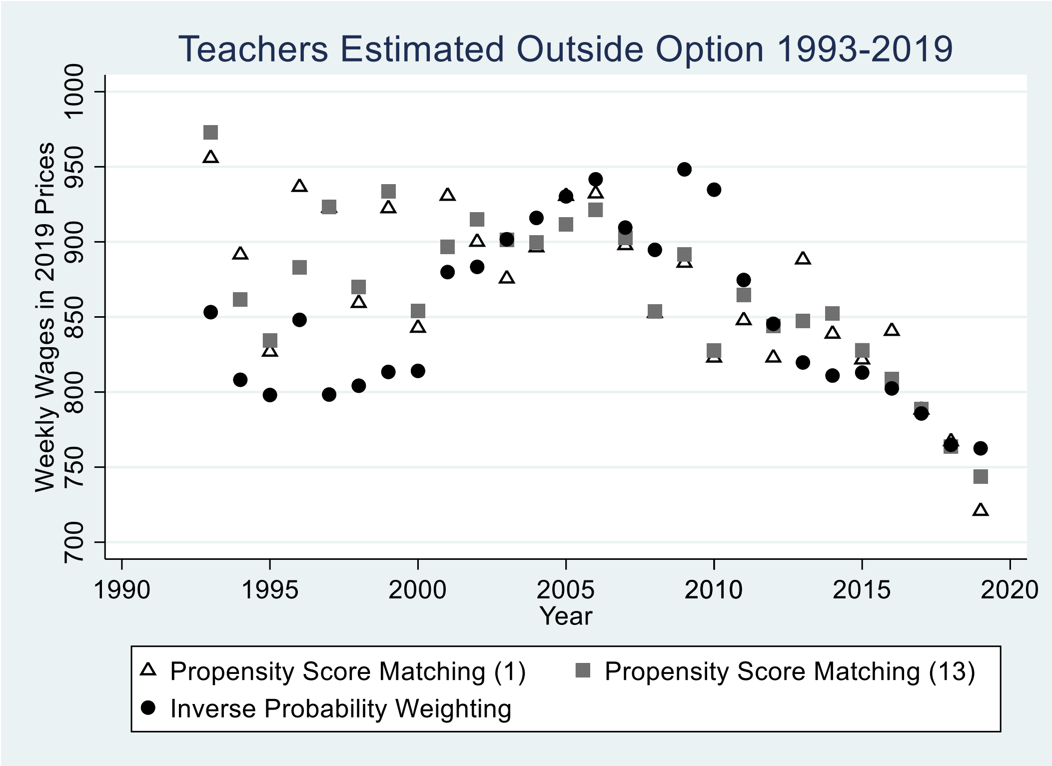
Our Dependent variables are standardized to have a mean of 0 and a standard deviation of 1. Regression includes all of our controls, these are: Class Size, Student Age above median in months, student sex, books at home, computer in home, speak English at home, teacher sex, teacher experience, teacher age and year fixed effects. Standard Errors in parentheses clustered at the classroom level while statistical significant is indicated by: +p<0.15,* p<0.10,**p<0.05,***p<0.01. Note: standard errors obtained from bootstrap (500)

Appendix Table 4 Effect of relative wages on Grade 8 pupil enjoyment in TIMSS

	1	2	3	4	5	6	7	8
	Science				Math			
Wage Difference (Match)	-0.0402 (0.0518)				0.0315 (0.0644)			
Wage Difference (Normal)		0.0204 (0.0960)				0.131 (0.105)		
Labor Market Returns to Teaching (Match)			-0.0430 (0.0518)				0.0469 (0.0652)	
Labor Market Returns to Teaching (Norm)				0.00885 (0.0935)				0.168+ (0.105)
Constant	0.680*** (0.0225)	0.680*** (0.0225)	0.680*** (0.0225)	0.680*** (0.0225)	0.607*** (0.0241)	0.606*** (0.0241)	0.607*** (0.0241)	0.606*** (0.0241)
DV mean (SD)	.745 (.435)	.745 (.435)	.745 (.435)	.745 (.435)	.640 (.480)	.640 (.480)	.640 (.480)	.640 (.480)
N	17156	17156	17156	17156	15060	15060	15060	15060

Regression includes all of our controls, these are: Class Size, Student Age above median in months, student sex, books at home, computer in home, speak English at home, teacher sex, teacher experience, teacher age and year fixed effects. Standard Errors in parentheses clustered at the classroom level while statistical significant is indicated by: +p<0.15,* p<0.10,**p<0.05,***p<0.01. . Note: standard errors obtained from bootstrap (500).

Appendix Figure 1 shows how the teachers' estimated outside option has changed using three different strategies. The black hollow triangle represents the estimates we use in this paper; these are calculated via nearest neighbour propensity score matching. The grey square is nearest neighbour propensity score matching, but slightly modified, as we increase the number of neighbours used to calculate the matched outcome to 13. Finally the filled black circle is teachers' outside option estimated using Inverse Probability Weighting.



Supplementary Material

Teachers Unemployment Rate

Specifically, we use the LFS to estimate the teachers' year and sex specific unemployment rate. This measure is the sum of unemployed individuals whose last job was teaching divided by the number of teachers plus the quantity of unemployed teachers. We estimate this separately by sex and year. Our measure of teacher unemployment only considers those who actually entered the teaching profession and therefore does not include those young people who want to go into teaching after they finished their training, but are unable to find a job. Although it is true that between 1 in 5 men and 1 in 10 women who finish teacher training do not to go into teachingthis does not mean that newly qualified teachers struggle to find a job as this is down to preferences and not employment opportunities. Each year roughly 3,000 more teachers leave the profession than enrol onto teacher training programmes. With pupil numbers increasing and more teachers leaving newly qualified teachers have extremely strong employment opportunities. Therefore any teacher unemployment we miss by using former teachers is unlikely to be significant. But if we measure teacher unemployment using qualified teachers we are likely picking up a lot of measurement error as many of these graduates may have never actually gone into teaching.

Teachers' unemployment rate tends to be around 1.7% and there are no meaningful gender differences. As the demand for teachers is driven by pupil numbers and policymakers desired pupil to teacher ratio we would not expect the teachers' unemployment rate to be affected by the financial crisis. However, we do observe that the unemployment rate rose above 2% between 2009 and 2012. We suspect this increase was driven by the fact that more than 50, mostly small rural Primary schools, closed during this period. It is important to note that the majority of the unemployment we observe among teachers is frictional as it is fairly unusual for teachers to get fired and the teachers who are affected by school closures tend to be amalgamated with another school. Similarly, we use the LFS to estimate the graduate unemployment rate by age, sex and year.

Teachers Relative Wages Descriptive Statistics using Merged Years

We have pupil performance data from the 1995, 2003, 2007, 2011 and 2015 TIMSS surveys. Additionally we are assigning each TIMSS teacher a teaching and non-teaching wage based on their sex (Male and Female) and age (measured in the following age bands: under 30, 30-

39, 40-49, 50-59 and 60+). To achieve the required sample size we merge the LFS years together in the following way: the TIMSS 1995 teachers wages are estimated using LFS data from 1993 to 1996, 2003 uses 2001 to 2004, 2007 uses 2005 to 2008, 2011 uses 2009 to 2012 and 2015 uses 2013-2017.

Consistent with our estimates from the previous section teachers tend to earn less than the average graduate but table 1a column a shows that when we account for non-random selection the difference falls significantly (from 17% to 8% and 13% to 7% in 1995 and 2015 respectively) or dissipates entirely (2003, 2007 and 2011). Male teachers face a significant pay penalty for remaining in the occupation (Table 1b) while female teachers have considerable pecuniary benefits (Table 1c).

Comparing earnings of current teachers to former teachers we have no strong evidence that teachers who quit the occupation sort into higher paying occupations (table 2a) however now that we have the power to split this by gender we find that, actually, male teachers sort into occupations that are 9% (2011) and 11% (2015) higher paying.

Table 1a Ratio of teacher and non-teacher wages using a matching strategy (normal strategy). Using the combined sample of men and women.

	(a)	(b)	(c)	(d)	(e)	(f)
			Age Group			
Times Year	All	U30	30-39	40-49	50-59	60+
1995	0.922	1.028	0.920	0.819	1.034	0.964
	(0.837)	(0.845)	(0.783)	(0.780)	(0.833)	(0.921)
2003	1.018	1.039	0.998	0.963	1.034	1.155
	(0.870)	(0.859)	(0.804)	(0.846)	(0.868)	(1.112)
2007	1.027	1.115	1.030	0.996	1.064	1.148
	(0.884)	(0.901)	(0.814)	(0.773)	(0.885)	(1.060)
2011	1.003	1.217	1.045	0.939	1.005	1.143
	(0.900)	(0.980)	(0.876)	(0.792)	(0.882)	(1.063)
2015	0.934	1.171	1.019	0.854	0.931	0.990
	(0.865)	(0.959)	(0.867)	(0.781)	(0.821)	(0.921)

Our matching strategy is estimating teachers' outside option using propensity score matching by matching teachers to non-teacher graduates who are working full time. The variables we match on are: ethnicity, sex, age, marital status and region. The normal strategy that is reported in brackets is simply the ratio of teacher and non-teacher mean earnings. All of these differences are significant to the usual levels unless specified.

Table 1b Ratio of teacher and non-teacher wages using a matching strategy (normal strategy). Using a sample of only males.

	(a)	(b)	(c)	(d)	(e)	(f)
			Age Group			
Times Year	All	U30	30-39	40-49	50-59	60+
1995	0.852	0.997	0.865	0.694	0.996	0.869
	(0.842)	(0.927)	(0.780)	(0.785)	(0.832)	(0.922)
2003	0.963	0.952	0.923	0.937	0.983	1.170
	(0.851)	(0.881)	(0.784)	(0.743)	(0.852)	(1.108)
2007	0.966	1.098	0.959	0.903	0.995	1.127
	(0.857)	(0.934)	(0.786)	(0.741)	(0.862)	(1.069)
2011	0.928	1.108	0.935	0.883	0.944	1.098
	(0.879)	(0.969)	(0.845)	(0.797)	(0.858)	(1.063)
2015	0.896	0.914	0.932	0.845	0.889	1.004
	(0.845)	(0.950)	(0.834)	(0.789)	(0.799)	(0.935)

Our matching strategy is estimating teachers' outside option using propensity score matching by matching teachers to non-teacher graduates who are working full time. The variables we match on are: ethnicity, sex, age, marital status and region. The normal strategy that is reported in brackets is simply the non-teacher mean earnings. All of these differences are significant to the usual levels unless specified.

Table 1c Ratio of teacher and non-teacher wages using a matching strategy (normal strategy). Using a sample of only females.

	(a)	(b)	(c)	(d)	(e)	(f)
			Age Group			
Times Year	All	U30	30-39	40-49	50-59	60+
1995	1.015	1.007	0.987	1.041	1.124	1.050
	(0.959)	(0.964)	(0.873)	(0.955)	(1.073)	(1.050)
2003	1.060	1.121	1.078	0.988	1.084	1.370
	(1.008)	(1.012)	(0.908)	(0.881)	(1.047)	(1.288)
2007	1.098	1.060	1.078	1.063	1.139	1.217
	(1.012)	(1.009)	(0.908)	(0.934)	(0.893)	(1.130)
2011	1.095	1.303	1.134	1.028	1.085	1.159
	(1.026)	(1.091)	(0.983)	(0.888)	(1.035)	(1.179)
2015	0.973	1.263	1.093	0.874	0.965	1.031
	(0.979)	(1.069)	(0.972)	(0.860)	(0.958)	(1.068)

Our matching strategy is estimating teachers' outside option using propensity score matching by matching teachers to non-teacher graduates who are working full time. The variables we match on are: ethnicity, sex, age, marital status and region. The normal strategy that is reported in brackets is simply the non-teacher mean earnings. All of these differences are significant to the usual levels unless specified.

Table 2a Ratio of teacher and non-teaching qualified teachers wages using matching strategy (normal strategy) using a combined sample of both men and women by age group by year

	(a)	(b)	(c)	(d)	(e)	(f)
			Age Group			
Times Year	All	U30	30-39	40-49	50-59	60+
1995	1.086	1.079	1.051	1.061	1.077	1.343
	(1.048)	(0.870)	(1.021)	(1.058)	(1.106)	(1.343)
2003	1.066	1.051	1.138	1.045	1.046	1.051
	(1.013)	(1.025)	(1.060)	(1.023)	(1.057)	(1.123)
2007	1.136	1.181	1.146	1.138	1.117	1.082
	(1.042)	(1.076)	(1.051)	(1.100)	(1.102)	(1.064)
2011	1.070	1.243	1.137	0.961	1.004	1.135
	(0.957)	(1.143)	(1.057)	(0.928)	(1.332)	(1.135)
2015	0.985	1.156	1.050	0.970	0.950	1.082
	(0.920)	(1.128)	(1.005)	(0.976)	(0.940)	(0.975)

Our matching strategy is estimating teachers' outside option using propensity score matching by matching teachers to non-teacher graduates who are qualified to teach and are working full time. The variables we match on are: ethnicity, sex, age, marital status and region. The normal strategy that is reported in brackets is simply the non-teacher mean earnings. All of these differences are significant to the usual levels unless specified

Table 2b Ratio of teacher and non-teaching qualified teachers wages using matching strategy (normal strategy) by sex and year

		Sex	
Times Year	All	Male	Female
1995	1.086	1.042	1.101
	(1.048)	(1.059)	(1.062)
2003	1.066	1.020	1.076
	(1.013)	(1.015)	(1.040)
2007	1.136	1.060	1.178
	(1.042)	(1.014)	(1.0789)
2011	1.070	1.004	1.094
	(0.957)	(0.916)	(1.003)
2015	0.985	0.914	1.022
	(0.920)	(0.893)	(0.956)

Our matching strategy is estimating teachers' outside option using propensity score matching by matching teachers to non-teacher graduates who are qualified to teach and are working full time. The variables we match on are: ethnicity, sex, age, marital status and region. The normal strategy that is reported in brackets is simply the non-teacher mean earnings. All of these differences are significant to the usual levels unless specified.

If only pecuniary factors matter, what quitting rates would we observe?

Teachers in England have a high rate of attrition, especially young teachers - according to the 2018 School Workforce Census (SWC), of the teachers who started in 2016 1 in 4 quit within 24 months. The relatively limited empirical evidence on the determinants of teacher attrition (Smithers and Robinson 2003, Stinebrickner 1998) suggests it should be modelled as some combination of pecuniary and non-pecuniary factors (as in Manski (1987)). Indeed a simple econometric model of occupational choice is that teacher i will continue to teach at time t if her expected utility for remaining in teaching (j) is greater than, or equal to, her expected utility in her next best non-teaching alternative (j'). Where her expected utility is some function of pecuniary (w) and non-pecuniary (g) job specific characteristics. Formally:

1. $$Teach_{it} = \begin{cases} 1 \; if \; EU\left(w_{ijt}, g_{ijt}\right) \geq EU(w_{ij't}, g_{ij't}) \\ 0 \; if \; EU\left(w_{ijt}, g_{ijt}\right) < EU(w_{ij't}, g_{ij't}) \end{cases}$$

Policymakers have largely focused on using pecuniary factors to reduce teacher attrition; recent policies include restructuring teacher training bursaries into early career payments and a commitment to increasing teachers' initial wages to £30k a year. As our estimates suggest that young teachers already tend to earn more in teaching than they would in their outside option, and enjoy higher job security, it seems unlikely that pecuniary factors motivate attrition. However, the growth in teachers' wages is typically slower than their outside option. As a consequence the decline in relative wages over the lifecycle might, partially, explain the high rates of attrition in England. In this section, we estimate the probability that, for a given age and sex, a teacher who leaves the occupation would maximise their lifetime earnings using the following logit model:

2. $$Pr(Y_{ays} = 1 \mid X) = \phi(\beta_0 + \beta_1 X_1 + \epsilon_{ays})$$

Y_{ays} is a dummy that indicates if for age a, in year y and for sex s the Net Present Value (NPV) for teaching is lower than the NPV of their outside option. We calculate the NPV of teachers and non-teachers using estimates obtained from the LFS. Specifically the teachers' wages are the mean earnings of all teachers in England for a given age, year and sex while their non-teaching wage is the average non-teaching graduates earnings, controlling for differences in observable characteristics via propensity score matching, for a given age, year and sex. $\boldsymbol{X_1}$is our vector of covariates, these are age (21-65), sex (Male vs Female) and year (1995, 2003, 2007, 2011, 2015).

To calculate the NPVs we assume that every teacher starts teaching at 21 and retires at 65 and their earnings over their lifecycle are the same as current teachers and non-teachers.[20] We are assuming that the unexplained component of teachers' wages is negatively correlated with the unexplained component of non-teachers' wages - teaching specific human capital is not rewarded on the labour market (Rickman and Parker 1990).

In addition, we are also assuming that there is no switching cost, a high (25%) or normal (12%) discount parameter, and that the market perfectly clears – they will be employed in teaching or non-teaching with a probability of 1.[21,22] Under these initial assumptions our estimates are intended to be interpreted as an upper bound.

Assuming a high (normal) discount parameter and perfect market clearance our logit estimates suggest that there is a 75% (77%) chance that male teachers could maximise their lifetime earnings by leaving teaching. While, consistent with the gender pay gap, we observe it is considerably less likely for female teachers (12% (9%)). The solid red line in figure 3 shows that the probability is highest for young teachers (88% (91%) for men and 21% (18%) for women) and lowest for those approaching retirement age (58% (57%) for men and 1%(3%) for women).

Relaxing our assumption on perfect market clearance and instead using the actual teacher and non-teacher unemployment rates we observe that the probability that a young teacher would be financially better off if they quit teaching falls - from 88% to 79% for men and 21% to 15% for women. As older graduates have a relatively low unemployment rate the impact of including employability on our estimates decreases with age to the extent that the probability for older teachers remains largely unchanged (see the green dot-dashed line vs the red solid line in Figure 3). If we impose a switching cost of 10% the probability does fall even more (from 75% to 60% for men and 12% to 6% for women), but even then there remains a high probability that young male teachers could maximise their lifetime earnings by quitting (see blue dashed line figure 3).

The probability that a male teacher would be financially better off if they left the profession exceeds 50% at almost every point over the lifecycle. Even if we assume a 40% switching cost, which is significantly larger than the impact of job displacement in our setting (Hijzen et al., 2010), we would still expect to observe an attrition rate of 33%. Yet, using the 2011 to 2018 SWC, we observe that

[20] For example in 2015 a 21 a female teacher earns £26kp.a, we will assume they will earn £34kp.a. when they turn 32, which is how much the average 32 year old female teacher earnt in 2015. We estimate the NPV separately by age (21-65), sex and year (1995, 2003, 2007, 2011, 2015).

[21] A discount parameter of 25% indicates that the value of getting £1 after one year and the £1 the year after has a net present value of £1.44 today (i.e. $\frac{1}{(1+0.25)^1} + \frac{1}{(1+0.25)^2} = 1.44$). While if we use a lower discount parameter (12%) the same income stream is worth £1.69 today (i.e. $\frac{1}{(1+0.12)^1} + \frac{1}{(1+0.12)^2} = 1.69$).

[22] Discounting rates tend to range between 10-14% (Meyer 2013) therefore we use the median (12%) as our normal discounting parameter. While our high discount rate is an arbitrary choice intended to show a scenario where individuals place a lot less significant on future earnings.

male teachers' actual rate of attrition is between 9.5-10.7%. This large discrepancy suggests that male teachers hold strong teaching specific non-pecuniary preferences and/or they are considerably misinformed about their outside option.

In contrast, for female teachers' the actual rate of attrition (9-10%) is consistent with what we would expect to observe if female teachers were trying to maximise their lifetime earnings (6-12%). As the labour market has become more female friendly it could be that the historic female specific non-pecuniary benefits to teaching (such as compatibility with household production and fertility choices) might not be as unique to the profession today as they once were. As a consequence, the attrition of female teachers could be, in part, driven by a desire to maximise expected earnings.