The Effects of Pay Decentralisation on Teachers' Pay and Teacher Retention

Jake Anders, Alex Bryson, Hedvig Horváth and Bilal Nasim*

NOT FOR GENERAL CIRCULATION

Abstract

In 2013/14 the arrangements for setting teacher pay in England were radically reformed. A system of seniority-based progression was replaced by a decentralised system allowing schools to pay teachers based on individual performance and local labour market conditions. Using a data-driven strategy, we classify schools into four groups according to what form of flexibility they adopted, then use this classification to analyse how pay flexibility affected teachers' pay and retention. About 9% of schools speeded up pay progression relative to what would have been expected in the seniority-based scheme, while about twice as many slowed it down. Teachers in these schools saw wages increase (decrease) by 4 (3) per cent on average. These wage changes seem to have little effect on teacher retention, suggesting school-specific labour supply in England is highly inelastic, implying large monopsony power for schools.

* University College London – Institute of Education. Anders: jake.anders@ucl.ac.uk, Bryson: a.bryson@ucl.ac.uk, Horváth: h.horvath@ucl.ac.uk, Nasim: b.nasim@ucl.ac.uk.

The authors would like to thank the Economic and Social Research Council for funding (grant number ES/R00367X/1); the Department for Education (especially Richard Howe and Gerard Hassett) for their assistance in arranging access to the Schools Workforce Census and providing invaluable advice on understanding unusual features of the data; and participants at the CEP Economics of Education Work-in-Progress seminar for helpful comments on an earlier version of this paper.

"You can't stop a teacher when they want to do something. They just do it."

— J.D. Salinger

1. Introduction

Good teachers matter, both in the short-run for pupil attainment, and in the longer-run for non-test outcomes (e.g. Chetty, Friedman and Rockoff 2014ab; Rothstein 2010,2015). Delivering good education therefore means attracting and retaining good teachers, and incentivising them to perform well. How to recruit, retain and incentivise good teachers is a concern in many countries. In England, concern has centred on the system of centralised pay determination since it is unable to respond to workers' outside options in local labour markets (Britton and Propper 2016). The School Teachers' Pay Review Body (STRB) which advises government on teacher pay awards in England noted in 2012 that "the current pay system is rigid, complex and difficult to navigate and does not support schools to recruit the high-quality teachers or leaders" (DfE 2012a).

This paper evaluates the impact of a major reform to that pay system which, for the first time, permitted local authority-controlled schools to pay teachers whatever they liked in order to attract, retain and incentivise them, subject to an upper and lower bound. We explore the effects of this shift away from the previous national seniority-based system towards one in which schools are free to negotiate pay with individual teachers as they see fit.

In principle, there is the potential for the reform to have significant effects not just on teachers themselves, but also their pupils, although it is the former on which we focus in this paper. Under the reform, schools are expected to have regard to individual teacher annual performance reviews. The previous spine-point pay schedule published by the STRB has been discontinued. However, teacher unions continue to provide 'shadow' national wage schedules intended to aid schools wishing to pursue the previous method of paying teachers. We use these to estimate the extent to which individual schools took advantage of the pay flexibility now available to them, explore the characteristics of the schools who chose to do so, and use a difference-in-differences strategy to estimate the effects of a school taking advantage of this flexibility on teachers' pay. We return to details of each of these aspects below.

The literature on rather small-scale incentive bonuses for teachers is large (e.g. Loyalka et al., forthcoming) but there are only a handful of papers about wage decentralisation on the whole teacher labour market, all of which are very recent. Biasi (2018) evaluates the impact of Wisconsin's Act 10 which permitted school districts to dispense with collective bargaining and replace it with individual wage bargaining. She finds districts that chose to switch experienced a greater pre-post improvement in value-added compared with districts that stuck with the regulated pay regime, which she decomposes into effects arising from effort (54 percent) and composition (46 percent), with the latter due to worker sorting by ability into districts with flexible pay schedules.

Willen (2018) investigates the wage effects of the cessation of centralised pay determination for teachers in Sweden with a system of individual wage bargaining devolving responsibility for pay setting to employers at local level. His setting differs from Biasi's (2018) in that schools and school districts do not get to choose whether or not to adopt decentralised pay setting: official centralised pay schedules were no longer published post-reform, meaning schools had to take some kind of active decision about pay setting going forward. He finds that the move to decentralised pay setting resulted in an across-the-board increase in teacher wages compared to the wages of college-educated non-teachers (CENTs) in the pre-reform period. This increase comes at the expense of other school expenditures, a resource switch which, Willen speculates, may partly explain the fact that the wage hike does not translate into an improvement in pupil attainment.

Sharpe et al. (2017) and Burgess, Greaves and Murphy (2018) examine the pay reform in England which is the focus of this paper. Sharpe et al. (2017) describe the evolution of teacher pay pre- and post-reform, while Burgess, Greaves and Murphy (2018) investigate the determinants of taking up flexibility. In contrast to the Swedish case they find teacher wages fall relative to what they might have been if centralised pay setting had remained in place. Furthermore, expenditure on teachers' wages constitutes a falling percentage of all schools' expenditure in the post-period. They interpret this as evidence that the centralised pay system had resulted in sub-optimal investment in teacher salaries, something schools were able to rectify with the shift to a flexible, decentralised pay regime. They also find that teacher salaries do not fall as much in local labour markets where teachers' outside labour market options are better paid and in areas where school competition is more fierce.

We revisit the English reform but ask a different question, namely what was the causal effect of the reform on teachers' pay?[1] We derive a measure capturing the extent to which schools vary teacher pay once they have the power to do so. This measure, which is similar to the dependent variable in Burgess et al. (2018), identifies the counterfactual pay progression that would have occurred for teachers in each school if the centralised system had remained in-tact. Teacher-level departures from those predictions are aggregated to school level to obtain an intensity-of-treatment measure for each school. Using this approach we classify schools into 4 groups namely: 1) non-adopters, who continued to pay teachers based on seniority as if reforms did not happen; (2) mean-zero adopters, who kept average pay progression as would have been expected based on the pre-reform pay scheme but increased within-school variation in pay progression; (3) positive adopters, who speeded up pay progression on average post-reform; and (4) negative adopters, who slowed down pay progression on average. We then use a difference-in-differences framework, treating non-adopters as the control group and adopters as various treatment groups, to estimate the effects of pay flexibilities on teacher pay and retention.

We choose the difference-in difference specification that does the most to control for unobserved pre-reform differences between the groups, while also relying on parallel pre-trends, which we confirm. About 9% of schools speeded up pay progression relative to what would have been expected in the seniority based scheme, while about twice as many slowed it down. Teachers in these schools saw wages increase (decrease) by 4 (3)% on average. These estimates imply, à la Manning (2003), very low labour supply elasticities to schools, leaving them with substantial monopsony power.

In the next version of this paper we will link our adopter classification to the Annual Survey of Hours and Earnings (ASHE), a one percent sample of all employees in employment in Britain – containing large samples of both teachers and non-teachers – that permits us to explore

[1] In future work we will also be examining the impact of the reform on other aspects of the teacher labour market, on schools' ability to recruit and retain teachers, and on pupil attainment.

the importance of local labour markets. We will also investigate the effects of the pay reform on the pay of incumbents and new entrants separately. This may very well turn out to be important since it may be easier for schools to adjust their wage bills via new entrants, rather than seeking downward adjustments in the wages of incumbents.

The remainder of this paper is set out as follows. Section 2 sets out the institutional background to the teacher pay reform in England. Section 3 describes the multiple sources of data that we use. In Section 4, we present our empirical strategy to classify schools according to their adopter status. Section 5 presents the difference-in-differences results before we conclude and lay out next steps in Section 6.

2 Institutional Background and the 2013-14 Teacher Pay Reform in England

In England, primary (age 5-11) and secondary (age 11-16) schools educate compulsory schooling aged children and young people.[2] About 94% of pupils study in the state-funded system, which comprises of local authority (LA) maintained schools (53% of pupils in state-funded schools are in LA maintained schools as of October 2018, DfE, 2019) and academies (47% of pupils). Schools in England have traditionally been funded through and subject to some level of oversight by their LA, typically at the level of either counties or unitary boroughs; this is intended to provide some level of democratic oversight and allow for area-level planning of matters such as supply of pupil places. However, the quality of this oversight in tackling chronic underperformance among some schools has been questioned by central government. In reforms starting in the early 2000s and accelerated post-2010, increasing numbers of schools are either forced (in the case of underperforming schools) or allowed (in the case of highly performing schools) to become

[2] Since 2015, young people have been required to take part in some form of education until they are 18. However, there are multiple pathways to do that: in full time education (for example remaining in a secondary school with a sixth form, at a standalone sixth form college, or at further education college), as part of an apprenticeship or traineeship; or receiving some education or training alongside 20 hours or more a week part time work or volunteering.

'academies' (Eyles and Machin, 2015). These schools are directly funded by the UK Department for Education (DfE) under an 'academy agreement', by which they receive slightly higher levels of funding but, in return, no longer receive services typically provided by their LA. They remain subject to some regulations (e.g. they are still inspected by England's educational inspectorate, Ofsted, on the same basis as LA-maintained schools) but the academy agreement allows increased freedoms in some respects, such as not having to follow England's national curriculum (but with a contractual agreement to ensure that their curriculum is broad and balanced).

Crucially, for the purposes of this paper, once a school switches from being an LA-maintained school to an academy by signing an agreement with the UK Department of Education, it has the freedom to set teacher pay as it sees fit. By contrast, up until school year 2012-13, classroom teachers in LA maintained schools were paid according to a rigid, essentially seniority-based pay schedule, which comprised of 2 pay scales. An example for that is shown in Panel A of Figure 1. New teachers entered the first spine point of the "main" pay scale and progressed through the 6 spine points year after year provided they completed a satisfactory year of employment. When reaching the top, they could apply for promotion to the "upper" pay scale which consists of 3 additional spine points. The schedules were centrally determined by the School Teachers' Review Body (STRB) and laid out as part of the "School teachers' pay and conditions document" issued annually by the DfE. These set the base pay of teachers. For additional responsibilities relating to teaching and learning, or to special educational needs, teachers can receive additional payments, which are subject to specific rules regarding the purpose and the monetary value of such payments (laid out in the same official document). There were, however, a number of avenues available to either head teachers or the relevant body to link pay to teacher performance, pre-reform (DfE 2012). However, as evidence in Annex B of DfE (2012) suggests, these were rarely used.

Two concerns arose with this pay regime. First, the system failed to reward teachers' performance. Figure 3 provides suggestive evidence for that by showing that in the period prior to the reform progression was nearly automatic on the first 5 spine points of the main pay scale, although it was slower higher up the scale. Second, the system did not respond to local labour market conditions which determine the attractiveness of teachers' outside options. As Figure 1 shows, apart from experience (spine points), only the very broad location of the school influenced teachers' base pay: whether they were in Inner-, Outer- or Fringe London or anywhere else in England ("Rest of England").

Hence, in 2012 the STRB outlined the case for reform of the pay system for teachers: "The current pay system is rigid, complex and difficult to navigate and does not support schools to recruit the high-quality teachers or leaders..." (DfE 2012a). To meet these challenges, the STRB recommended increased flexibility in determining teacher salaries and linking teacher pay to performance, with higher rewards and more rapid progression for the most able teachers (DfE 2013). The government accepted the key recommendations in full, and the reforms came into effect in 2013 for new teachers and for all in 2014 (STRB 2013).

They involved the introduction of additional flexibilities and greater encouragement of their usage. The most visible aspect of the reform was that the STRB only prescribed a maximum-minimum pay range for the main and the upper pay scale but ceased to designate pay points in between. Panel B of Figure 1 illustrates the change. In addition, STRB made the following recommendations:

- Progression through the main pay scale was to be directly linked to excellence and performance improvement, replacing increments based on length of service with pay progression linked to annual appraisal for all teachers.
- Mandatory pay points were abolished within pay scales for classroom teachers with the intention to allow pay decisions to be made at the individual level.
- Related reforms allowed schools the flexibility to create posts paying salaries above the upper pay scale and reduced restrictions on schools in the use of allowances for recruitment and retention.

However, little further guidance was offered to schools by the DfE on what kind of changes to implement.[3] This was by design, the DfE arguing that "It is up to each school to decide for itself how best to implement the arrangements and develop its pay and appraisal policies accordingly"

[3] An optional model appraisal policy was issued in 2012 (updated in 2019; DfE, 2019) but this simply sets out the process for schools making a recommendation on pay, which could take any form depending on how a school has decided to organize its pay system.

(DfE, 2019). Thus, while the reforms offered an *opportunity* for schools to flex their pay, there is no specific mandate to do so other than the general guidance “pay decisions must be clearly attributable to the performance of the teacher in question” (DfE, 2019), which others have interpreted continue “to permit the adoption of fixed pay scale points as the basis for pay progression” (Unions, 2018). Indeed, a survey of teachers in 2015 suggested that a third believed performance was not being used as the basis for pay decisions in their schools (Busby, 2015). This poses a methodological challenge when evaluating the impacts of the reform, which we will overcome in a data-driven way.

Teacher unions were concerned that, faced with budget cuts, schools will use their new powers of flexibility to freeze pay, possibly to the detriment of women and minorities (NEU, 2019). Therefore, unions continue to designate and publish reference points within the maximum and minimum pay range prescribed by the STRB. Figure 2 illustrates these recommendations. In the accompanying documentation unions remind schools that retaining a “fixed” (seniority based) pay ladder is permissible and that the unions’ intention is to assist in this regard by continuing to publish reference points (Unions 2018). This unique feature of the pay reforms essentially allows us to observe counterfactual teacher pay after the reforms: we rely on this in our empirical strategy when classifying schools into adopters vs. non-adopters in a data-driven way. We turn to this procedure in Section 4, after a brief description of the data.

3 Data

In addressing the effects of pay reform on teacher pay we use teacher- and school-level records from the School Workforce Census (SWF) linked to the School Performance Tables, Consistent Financial Reporting returns, and the Annual Survey of Hours and Earnings. We describe each of these in more detail next, as well as the linkages between them that we exploit.

3.1 School Workforce Census (SWF)

The School Workforce Census (SWF) provides data on all school workforce employees including teachers, teaching assistants, other support staff and auxiliary staff, working in publicly funded schools in England. It is an annual, individual-level, employer-collected, mandatory administrative database administered by the Department for Education (DfE).[4] It covers demographic and job-level characteristics such as gender, age, ethnicity, disability, grade, job role, and qualifications. For teachers, it also includes their pay scale, salary, weekly hours worked and sickness absence data. Subject specialism is available for secondary school teachers. Up to now, data through 2010 to 2016 have been made accessible for research, and the longitudinal dataset includes unique teacher identifiers to track teachers within and across schools over time. The SWF contains school-level information on teacher vacancies and includes school identifiers which can be used to link to other datasets such as the School Performance Tables.

As the reform intended to offer more flexibilities to LA maintained schools, our base sample consists of schools that were LA maintained in 2012, just before the reforms kicked in (but may have converted to academies later). We exclude schools that merged/split, but these represent a small fraction (<1%).[5,6] To classify schools into groups of adopters/non-adopters, we use contracted, nondaily rate classroom teachers on the main or upper pay scales working in the base sample of schools. When analysing the effects of adopting pay flexibility, we use all contracted, non-daily rate classroom teachers who work in one of the base sample schools in any given year. Summary statistics for the base sample of schools and teachers are shown in Table 1 and 2. Primary schools are overrepresented in our sample because fewer primary schools academised. In addition to school and teacher identifiers, we use data on the pay scale and spine point of teachers before the reforms, average weekly hours described by full-time equivalency

[4] https://www.gov.uk/guidance/school-workforce-census

[5] We exclude schools who were academies before the reform simply because they were not intended to be treated in any way. Schools with a strong latent preference for pay flexibility might academise earlier in order to flex their pay, so those who remained LA maintained throughout the pre-reform period are likely more similar in this regard.

[6] We define schools by their unique reference number (URN), which is generally stable over time but can change in the event of a major reorganization of the school such as academy conversion or merger with another school. However, as there can be circumstances where this identifier does change but we wish to ensure we track the school despite this, we make use of the Consistent Schools Database (https://www.closer.ac.uk/work-package-5-csd) created by a member of the project team which provides a cross-reference for changes in this identifier.

status, full-time equivalent base pay and gross pay. Overall pay consists of both base pay – the minimum and maximum values of which are regulated by the STRB - plus any additional pay components received for extra responsibilities such as teaching and learning responsibility (TLR) or special education (SEN) allowance. Yearly summary statistics on the pay variables show that in constant 2015 sterling pounds pay decreased in the pre-reform period and has stagnated since.

There are three data issues that affect our work. First and most importantly, how schools report pay to DfE varies by school and over time, in ways that may correlate with schools adopting flexibility. Up until 2013, schools mandatorily reported pay scale and spine points, through which DfE linked base pay mechanically. Therefore, supposedly, there is little measurement error in pay variables up to 2012. In 2013, DfE preferred if schools also report base pay themselves, beyond pay scale and spine points, which many schools did.[7] From 2014 onwards, schools were required to report base pay itself, as mandatory spine points were discontinued. In addition, schools use different software programmes to file their reports with DfE, some being better than others, so this introduces school-level measurement error of different magnitudes. DfE do have flag variables for software use: when we have acquired these data, we will use them to investigate measurement error. In the meantime, we bound the magnitude of measurement error by comparing all results to 2013 results, which year we exclude from both the pre-reform and the post-reform period. Recall that in this year schools reported both spine points and base pay, however, the quality of reporting was exceptionally low (see footnote 7). Therefore, DfE in personal communication suggested to avoid this year of data if possible. In addition, 2013 was an intermezzo in the sense that the reform then took effect for new teachers but not for existing ones. All this backs up our choice to treat 2012 as the last (clean) pre-reform year and 2014 as the first (clean) post-reform year, and we will use pre- vs. post-reform terminology accordingly

[7] Potentially due to this issue the quality of pay data is of much lower quality in 2013 than in other years. This is apparent from the standard deviation of raw pay variables, which are 35 times larger than in previous years, and 15 times larger than in subsequent years. Personal communication with DfE confirms our conjecture that this was due to some schools reporting full-time equivalent base pay in lieu of raw base pay, which led to double full-time adjustment in these cases. Obvious cases are easily corrected by relinking raw base pay through pay scale and spine point variables. The rest is taken care of by a 1%, two-sided winsorising of the pay variables for every region, pay scale and spine point variable, which we do for all other years as well. In post-reform years, we perform the winsorising using counterfactual spine points. See Section 4 for more details.

through the rest of the paper. Nevertheless, we report results for 2013 for completeness, we do not draw conclusions from it. We may interpret it as an upper bound of all measurement error, however, it may also reflect some immediate impacts of the reform (see Section 5).

Second, all pay variables are available rounded to the nearest 1,000 only. This introduces measurement error into our analysis. In pre-reform years, when schools report pay scale and spine points through which DfE mechanically links base pay, we can recover exact pay. This allows us to explore patterns of measurement error due to rounding, and it appears fully classical.

Finally, the number of teachers we see per school (and year) may be small. We plan to refine our empirical strategy to statistically account for this.

3.2 Consistent Financial Reporting (CFR)

Consistent Financial Reporting (CFR) returns provide data on LA maintained schools' per pupil income and expenditures. Income is broken down into 2 categories: self-generated vs. grant income, while expenditures are split into multiple categories, including expenditures on teachers, supply teachers or educational support staff. We link these data to the SWF at the school level; summary statistics are shown in the lower panel of Table 2.

3.3 School Performance Tables

Annual school-level aggregates of pupils' performance at various points of their educational careers are publicly available from the DfE website. These are aggregated from pupil-level performance data held within the DfE's National Pupil Database (NPD) and are primarily published for the purpose of providing parents with information to inform school choice decisions. There are legal limitations on the linking of pupil-level and teacher-level data. However, we are able to link these published school-level aggregates with the school-level variables we have aggregated from the SWF using administrative school identifiers (UPNs) published by the DfE. In most cases these aggregates are simple means constructed from all pupils in the relevant year cohort in a school. However, some sub-groups are available, specifically splitting individuals into 'low', 'average' and 'high' prior attainment, and identifying average performance among those pupils who are registered as eligible for free school meals (a proxy for individuals in low income households). The published measures are based on tests taken by virtually all pupils in state funded schools at age 11 (end of primary school; linked to our school-aggregate teacher pay measures for primary schools) and age 16 (end of secondary school; linked to our school-aggregate teacher pay measures for secondary schools). The end of primary school tests are low-stakes for children, but potentially have accountability consequences for the schools. The end of secondary school tests are high-stakes, as these are pupils' 'GCSE' examinations, which are a key qualification for progression to further and higher education in the English context. Both raw average pupil achievement and 'value added' measures are available for each of these stages. The value-added measure for primary schools is based on the difference between their performance at age 11 and assessments from early in their primary school careers. The value-added measure for secondary schools is based on the difference between their age 16 performance and their age 11 performance (the same tests that form the outcome measures for primary schools).

3.4 Annual Survey of Hours and Earnings (ASHE)

The Annual Survey of Hours and Earnings (ASHE) is a linked employer-employee panel data set containing a 1% snapshot sample of all workers in employment. The employers provide a wide range of information about the employee's hours and earnings during the preceding year, including the amount of bonus or incentive pay received during the current pay period. It also contains both residency and workplace postcodes. The survey is carried out by the UK's Office for National Statistics (ONS) and is mandatory. Currently, we only use publicly available, local authority level information about mean wages from these data, but in future we will be using this dataset when we add the local labour market component of the analysis. We can identify teachers and schools in the dataset and merge in school-level information from SWF including the variable we construct to classify schools into adopters/non-adopters. Hence, we will be using ASHE extensively to construct a travel distance based local labour market for teachers and estimate their outside options, to identify new entrants to the profession and to track pay progression of teachers over time. ASHE is available from 2002, so from much earlier than SWF, and this will also help us to ensure parallel pre-trend assumptions for our DiD analysis.

4 Have schools been using the flexibilities, which schools and how?

As described in Section 2, a unique feature of our setting is that even after the reform, a consortium of large teacher unions has continued to designate and publish reference pay points within the binding maximum-minimum pay range. This allows us to construct counterfactual pay for every teacher, had the pay reform not taken place. We use these counterfactual salaries to identify schools that adopted some form of flexibility and those who continued to use the seniority-based system. Hence, we classify schools into adopters and non-adopters. We also distinguish between groups of adopters according to the form of flexibility they adopted. This section describes step by step our classification procedure to do that. In doing so we address whether schools have used the flexibilities they were offered, which schools have done so and what strategies they adopted.

4.1 Methods

4.1.1 Constructing counterfactual pay

We start by using spine point data from years 2010 to 2013 to predict counterfactual progression through the pay ladder had the reforms not taken place. We combine two pieces of information in pre-reform data to do this: (i) average individual-level progression within the main or the upper pay scale, and (ii) average progression from given spine points in the school (or in the LA when there is insufficient number of observations within the school). That is, we estimate the following regression for years $t = 2011{,}2012{,}2013$:

$$\Delta spine\ point_{it} = \sum_{s} \sum_{p=1}^{9} \beta_{sp} \times W_{i,t}^{s} \times D_{i,t-1}^{p} + \varepsilon_{it},$$

where $\Delta spine\ point_{it}$ is the change in spine points of teacher *i* between years *t-1* and *t*, and spine points run from 1 to 9, 1 to 6 corresponding to the main pay scale spine points, while 7 to

9 correspond to the upper pay scale spine points. $W_{i,t}^{s}$ is a set of dummy variables indicating whether teacher *i* in year *t* works in school *s*, while $D_{i,t-1}^{p}$ ($p = 1,2,\dots,9$) is another set of dummy variables indicating whether teacher *i* in year *t-1* was on spine point *p*. β_{sp} is the school- and lagged spine point-specific coefficient for the interaction terms. $\hat{\beta}_{sp}$ will be used to estimate average progression from given spine points in the school.

The within-teacher average of the residuals, $\hat{\varepsilon}_{it}$, across the years when the given teacher moved within the main pay scale, $\frac{1}{M}\sum_{\tau:\, 1 \le p(i,\tau-1) \le 5} \hat{\varepsilon}_{i\tau}$, or the upper pay scale, $\frac{1}{U}\sum_{\tau:\, 6 \le p(i,\tau-1) \le 9} \hat{\varepsilon}_{i\tau}$, will be used to estimate individual-level progression, where *M* is the number of years teacher *i* was on the main pay scale, and where *U* is the number of years teacher *i* was on the upper pay scale. Therefore, predicted progression for teacher *i* in in school *s* in years *t=2014,2015,2016* is

$$\widehat{progression}_{it} = \begin{cases} \hat{\beta}_{sp} + \dfrac{1}{M} \displaystyle\sum_{\tau:\, 1 \le p(i,\tau-1) \le 5} \hat{\varepsilon}_{i\tau} & \text{if } p(i,t-1) \le 5 \\ \hat{\beta}_{sp} + \dfrac{1}{U} \displaystyle\sum_{\tau:\, 6 \le p(i,\tau-1) \le 9} \hat{\varepsilon}_{i\tau} & \text{if } p(i,t-1) \ge 6 \end{cases}$$

Whenever the individual component cannot be estimated (due to the given teacher not seen on one or the other pay scale) progression is estimated with the school-level component only. Counterfactual spine points then for teacher *i* in years *t=2014,2015,2016* are computed as the sum of spine point last year and predicted progression, rounded to the nearest integer:

$$cf\ spine\ point_{it} = int\left(spine\ point_{i,t-1} + \widehat{progression}_{it}\right).$$

This procedure gives us counterfactual spine points for 92.8%, 88.5% and 83.6% of the pay sample in years 2014-2016.[8] The ratio is declining as we do the prediction further out in time after the reforms kicked in due to the cumulative nature of the procedure.

Once we have the counterfactual spine points, we merge in associated base pay from the reference pay points designated and published by the consortium of large teacher unions. As a result, we have both observed and scheduled (full-time equivalent) base pay for every teacher, where scheduled pay is conceptualised as counterfactual pay had the reforms not taken place.

[8] Non-missing rates for spine points in the pre-reform years vary between 98.2% and 99.4%.

We use both next to figure out in a data-driven way which schools departed from the seniority based regime and which schools did not.

4.1.2 Have schools been using the flexibilities and how? -- Classification of schools

The idea behind our classification procedure is that in schools that did not move away from the seniority based pay regime, scheduled full-time equivalent base pay ("scheduled pay," computed as described in subsection 4.1.1, rounded to the nearest 1,000, and deflated to constant 2015 £s) should perfectly predict observed full-time equivalent base pay ("observed pay," which is only available for researchers rounded to the nearest 1,000, and is deflated to constant 2015 £s). In contrast, in schools that adopted some form of flexibility scheduled pay's predictive power should be lower. We formalize this idea by computing the root mean squared error (RMSE) of observed pay when predicted by scheduled pay in every school for the post-reform period (2014-16):

$$RMSE_s = \sqrt{\frac{\sum_{i,t \in \{2014,2015,2016\}\ s.t.\, s(i,t)=s} (observed\ pay_{it} - scheduled\ pay_{it})^2}{n_s}},$$

where n_s is the number of teacher-year observations in school s. We can interpret RMSE as the average deviation of observed pay from scheduled pay in a school. It is essentially the quadratic mean of teacher-year-level deviations, which treats negative and positive deviations symmetrically. In case of perfect fit, RMSE is zero. We compute the same object for the pre-reform period (2010-2012), and for each year separately as well.

Figure 5 shows the histogram of RMSEs pre and post-reform. Before the reform, between 2010-12, school-level RMSEs were very small. The stark difference in the pre- and post-reform distribution provides evidence that schools have used the flexibilities they were offered and have departed from the seniority based pay regime. Pre-reform, 89.6% of schools had a perfect fit of observed pay on scheduled pay with RMSE exactly 0, while post-reform there are only 277 such schools, constituting 1.7% of all schools in the sample. While the most conservative classification should treat only these 1.7% as non-adopters, we are more lenient, since we know there is measurement error both in the observed pay data and, due to having predicted them, in the scheduled pay data as well. Therefore, we classify schools into adopters vs. non-adopters based on a cutoff value in their post-reform RMSE. We choose this cutoff, somewhat arbitrarily, to be £800, but we will see some justification for this choice below in Figure 6. Also, in future work, we will perform sensitivity analysis with regards to it.[9] Such leniency in the classification most likely leads to a clean adopter group, however, the non-adopter group may contain some misclassified adopter schools. Eventually, the classification results in 13,869 adopter schools and 2,348 (15%) non-adopters.

Panel A-B in Figure 6 displays some diagnostics regarding the classification. Panel A shows the RMSE over time in adopter schools and non-adopter schools. By construction, adopter schools have higher RMSE post-reform. It is notable, however, that pre-reform the two groups are identical in terms of RMSE. In 2013, RMSE in both groups jumps up – as explained above, this provides an upper bound on the size of the measurement error in our setting. What is reassuring is that RMSE in non-adopter schools remain stable at this level – and this provides the justification for our choice of the adopter/non-adopter RMSE cut-off value.

What strategies have adopter schools employed to use their flexibilities? To shed light on this we look at the school-level mean deviations of observed pay from scheduled pay. This is shown over time in Panel B of Figure 6. The graph shows no difference at all between adopters and non-adopters. However, as schools were given no guidance about what strategies to employ, the average may mask important differences in what they actually did. Therefore, we refine the classification and partition adopter schools into 3 groups:

1. "mean-zero adopters" are adopter schools where pay progression on average continued just as expected based on the pre-reform, seniority based schedule, but they increased within-school pay progression differences across teachers. In these schools RMSE increased relative to that of non-adopter schools but mean deviation of observed pay from scheduled pay remained 0. Beyond the RMSE criterion above, we select schools

[9] We also plan to treat these RMSE values as continuous measures of the degree of adopting flexibility, as well as to refine this approach to account for the small number of teachers in a non-negligible number of schools using shrinkage methods.

whose mean deviations of observed pay from scheduled pay fell between -800 and 800. 9,506 (58%) schools fall into this category.

2. “positive adopters” are adopter schools where pay progression on average speeded up relative to what was expected based on the pre-reform seniority based schedule. In these schools both RMSE and mean deviation of observed pay from scheduled pay increased relative to their level in non-adopter schools. Beyond the RMSE criterion above, we select schools whose mean deviations of observed pay from scheduled pay exceed 800. 1,421 (9%) schools fall into this category.

3. “negative adopters” are adopter schools where pay progression on average slowed down relative to what was expected based on the pre-reform, seniority based schedule. In these schools RMSE increased and mean deviation of observed pay from scheduled pay decreased relative to their level in non-adopter schools. Beyond the RMSE criterion above, we select schools whose mean deviations of observed pay from scheduled pay remains below -800. 2,842 (18%) schools fall into this category.

That is, we do the partitioning by applying another cut-off rule but now on the mean deviation of observed pay from scheduled pay.[10] Table 3 sums up the resulting number of schools by phase (primary/secondary) and adopter status.

Panels C and D in Figure 6 show the evolution over time of RMSE and mean deviation between observed and scheduled pay, by the more elaborate adopter status. The graphs illustrate very well each school type’s strategy. Mean-zero adopters departing from scheduled pay but only by progressing some teachers faster, while other slower, and these offset each other. Positive adopters exceed scheduled pay, while negative adopters fall behind of it. We can also illustrate these strategies by directly plotting observed pay against scheduled pay for the different adopter types, pre- and post-reform. This is done in Figure 7, which shows the empirical pay schedule, the non-parametric empirical relationship between observed pay and

[10] We choose this cutoff to be -800 and 800, of the same absolute value as the RMSE cutoff, on purpose. As mentioned above, RMSE is the quadratic mean deviation of observed pay from scheduled pay, while the mean deviation is obviously the arithmetic mean of the same. By the inequality between the two, RMSE is larger or equal than the absolute value of the mean deviation. Hence the choice of the same cutoff.

scheduled pay. Panels A-B show that pre-reform, all schools' pays line up on the 45-degree line, where observed and scheduled pay are equal. However, post-reform, only non-adopters fit right on the 45-degree line, while adopter schools deviate from it. Panel D also reveals that positive adopters are most generous with (so depart positively the most for) otherwise low-paid teachers, while negative adopters depart most in the middle and at the very top. This figure brings us closer to understand better what schools actually did when they had the option to flex their pay, and we intend to dig deeper in this in future work with the aim to identify typical school practices.

Having classified schools into the 4 groups above, we use this classification in Section 5, in our difference-in-difference estimation. But let us first see what determines a school's adopter status. It is an interesting prediction exercise on its own but is important even more so, as we will need to control for such determinants in our efforts to ensure the parallel pre-trends assumptions of the difference-in-difference approach.

4.2 Which schools are the adopters?

To see which schools are the adopters, we do a model selection: starting from an OLS regressions of adopter status indicators on different school, teacher, pupil and local labour market characteristics, we select the relevant predictors using AIC criterion. In each model, the dependent variable is 1 for the given adopter indicator and 0 for non-adopters only.[11]

Results are shown in Tables 4. On the one hand, we can see that schools are more likely to adopt some form of flexibility if they are

- within London;
- larger in size (in terms of teacher numbers);
- have a young teacher workforce;
- have more pupils eligible for free school meals;

[11] A more elaborate prediction exercise will be carried out, which allows for arbitrary interactions between the different predictors. This is important as it seems that some predictors (e.g. share of academies in LA) are more important for secondary schools, for instance, than for primary schools, however, these are masked here due to the relatively small share of secondary schools in the sample.

- experience higher teacher turnover.

The last two are economically small effects. On the other hand, schools that have a larger pay bill and spend a larger share of their budget on teachers are less likely to adopt any kind of flexibility, especially the one that speeds up pay progression.

It is also apparent that local authorities play an important role in determining adopting status: whenever LA effects are included, the predictive power (measured by R-squared) of the regressions goes up. This is not surprising given that LAs may publish model pay policies that schools can readily adopt if they wish. We are in the process of hand-collecting these model pay policies from each LA and year, which will allow us to cross-validate our school classification into adopter categories.

This prediction exercise is an important input for the rest of the paper. In section 5, we analyse the effects of pay flexibility on teacher pay, hours, retention and exit from the LA-maintained school sector in a difference-in-difference framework, using the adopter status as treatment. Therefore, if there are school characteristics that do predict adopter status, at least we will have to control for them to approximate conditional independence. We discuss our approach in more detail next.

5 The effects of adopting flexibility – A difference-in-differences analysis

Having identified schools that adopted flexible pay, in this section we turn to analysing the effects of adopting this pay flexibility. We focus on the following outcome variables: teacher pay, hours, retention and exit from the LA-maintained school sector, but we plan to expand this list to see if teachers became more likely to move across certain types of schools, if certain types of schools' ability to fill vacancies changed, or if the teacher workforce composition changed. We employ a difference-in-difference (DiD) framework, using the adopter status as a treatment indicator; we discuss the challenges we face in using adopter status in this way.

5.1 Empirical specification

We estimate the following model:

$$y_{it} = \alpha_0 + \sum_{\tau \neq 2012} \alpha_\tau \times year_\tau + \beta \times adopter_{s(i)} + \sum_{\tau \neq 2012} \gamma_\tau \times year_\tau \times adopter_{s(i)} + controls_{it} + \varepsilon_{it}$$

where

- y_{it} is outcome of teacher *i* in year *t*: log of full-time equivalent base or gross pay (in constant 2015 £s), full-time equivalent hours and probability of retention at same school;[12]
- $year_\tau$ are year dummies, with 2012 being the reference year (last pre-reform year);
- $adopter_{s(i)}$ is the adopter status of teacher *i*'s school, *s* ("mean zero", "positive" or "negative" adopter), always compared to non-adopters only;
- $controls_{it}$ include school effects and teacher demographic characteristics (age, age squared, female indicator and white British indicator) and pre-reform pay point.

The coefficients of interest are γ_τ's (the DiD estimates), and we calculate standard errors clustered at the LA-level, which we view as conservative.

5.2 Identification assumption and adopter status as a treatment variable

The identification assumption behind our empirical strategy is the familiar difference in differences assumption of common trends between treatment and comparison groups in the absence of treatment. In this case, this means that, in absence of the reform, each of our adopter groups follow the same time path in our outcome variables of interest as the non-adopters. This is, to some extent, mechanically true for our immediate pay outcomes in that all schools would have continued to be constrained to follow the seniority pay structure previously described. However, there are limitations to this argument. First, some schools with a strong desire to adopt

[12] We plan to add more outcome variables such as probability of teachers moving between certain types of schools, schools' ability to fill vacancies, or teachers' different observed characteristics to see if the teacher workforce composition changed.

pay flexibility could choose to become academies as an alternative route to gaining this flexibility; we currently exclude this possibility by restricting our sample to only schools that do not become academies throughout our period of analysis, so that academy conversion as a major reorganization does not confound pay reform effects. Furthermore, similar pre-trends in pay when it has been constrained by policy are likely to have disguised latent preferences for pay flexibility that the reform has allowed us to observe. To the extent that such latent preferences are correlated with trends in our other outcome variables of interest, this could pose a threat to our identification of causal effects. Where such correlations are explained by observable school characteristics, inclusion of these as covariates should help to mitigate this bias; however, clearly there may well be unobservable characteristics that we are unable to control for in our modelling.

As such, we examine the plausibility of the common trends assumption by visual exploration of pre-reform trends in the outcome variables of interest. In future work, after linking our school-level indicator of adopter status to the ASHE dataset, we will be able to observe and analyse much longer pre-trends (going back until 2002) to provide additional evidence to support this identification assumption. For the time being, these trends go back to 2010.

5.3 Results and Discussion

Figures 8-11 and Tables 5-8 display the results for different outcomes. In almost all of the figures, pre-trends, or rather the lack of them, are reassuring regarding the validity of the design.[13]

Figure 8 (and associated Table 5) shows the "first stage" of the reform in a sense, by illustrating how full-time equivalent base pay was affected by adopting flexibility. We see that overall, adopting some form of flexibility slightly increased base pay, however this small effect masks large heterogeneities in the form of flexibility. The mean-zero strategy does hardly anything to base pay, while positive adopters saw their base pay increase by 4%. The strategy that slows down pay progression relative to what would have been expected under the old, seniority based regime ("negative adopters") results in base pay falling by 3%. As shown by Figure

[13] The only exception is Panel D of Figure 11, which suggests that negative adopters had had higher (although converging) retention rates.

9 (and Table 6), results for gross pay show very similar patterns, suggesting that schools did not adjust payments for additional responsibilities (or the additional responsibilities themselves) to compensate for gains/losses in base pay.

The pay variables above were computed to be full-time equivalent, so if any adjustment in hours took place in response to the reforms, it would not show up in those results. Therefore, we separately analyse full-time equivalent hours. Figure 10 (and Table 7) show the results that the reforms, no matter what form of flexibility a school adopted, did not bring about any changes at all in hours as a margin of adjustment.

Next, we look at whether adopting any of the three forms of flexibility was successful in retaining teachers at the same school. Therefore, in these specifications we run linear probability models with the dependent variable of observing a teacher in the same school as they were observed the previous year on the left-hand side. Figure 11 (along with Table 8) display results, which are not dramatic. Overall there are no statistically significant effects of all three forms of flexibility jointly on retention. Point estimates suggest that speeding up pay progression relative to what would have been expected under the seniority based regime ("positive adopters") has the potential to increase retention. However, all these, as well as the other effects are noisy.

From these estimates, we can recover the implied school-specific labour supply elasticity as follows (Manning 2003):

$$\epsilon^{ssls} = -2\frac{-\gamma^{DiD}_{retention} / separation\ rate}{\gamma^{DiD}_{\log pay}} = \frac{2\gamma^{DiD}_{retention} / separation\ rate}{\gamma^{DiD}_{\log pay}},$$

which is small when plugging in our numbers for positive adopters:

$$\epsilon^{ssls} = \frac{2\gamma^{DiD}_{retention} / separation\ rate}{\gamma^{DiD}_{\log pay}} = \frac{0.02/0.2}{0.04} = 2.5.$$

This estimate is remarkably in the same ballpark as that of Ransom and Sims (2010), meaning that school-specific labour supply is relatively inelastic, implying strong monopsony power for schools.

6 Conclusion and Next Steps

In this paper, we analysed the impacts of the 2013-14 teacher pay reform in England, which scraped the seniority based pay ladder for teachers, leaving only a binding minimum-maximum range, within which schools are free to set pay. The reform was implemented to enable schools to respond to the local labour market conditions of their area, and to be able to attract and retain effective teachers.

The reforms did not give any guidance to schools on how to set their wages post-reform, so schools were only given an opportunity. To figure out which schools used these newly earned freedom, we develop a data-driven approach to classify schools into 4 groups: (1) non-adopters, who continued to pay teachers based on seniority as if reforms did not happen; (2) mean-zero adopters, who kept average pay progression as would have been expected based on the pre-reform pay scheme but increased within-school variation in pay progression; (3) positive adopters, who speeded up pay progression on average post-reform; and (4) negative adopters, who slowed down pay progression on average. The classification leaves us with about 15% of schools falling into non-adopters, 9-18% of positive- and negative adopters, and the rest ending up in mean-zero adopters.

Next, we use a difference-in-difference framework to estimate the effect of pay reforms on various outcomes such as teacher pay, hours and retention, using adopters as treatment groups. Overall we find a negligible increase in teacher pay, with a much larger (4%) rise at positive adopter schools, and a 3% cut in negative adopter schools. We find no effects on retention. These estimates imply very low labour supply elasticities to schools, leaving them with substantial monopsony power.

In near-future work, we plan to extend this paper in several aspects. First, enrich the DiD analysis with heterogeneity analysis and further outcomes, such as teacher entry wages, teacher mobility between the different types of schools, teacher workforce composition within school, or schools' ability to fill vacancies. Second, we intend to refine the classification procedure of schools into adopter groups along 3 dimensions: (i) by taking into account the fact that some schools have a small number of teachers, which makes their classification less accurate; (ii) by doing sensitivity checks on the classification cut-offs; (iii) allowing for differences in the timing when schools adopt pay flexibilities and switch to an event study design. We will also be able to cross-check our classification when we have collected all pay policy documents from Las.

Third and most importantly, we extend the analysis by bringing in the Annual Survey of Hours and Earnings (ASHE), a linked employer-employee panel data set containing a 1% snapshot sample of all workers in employment, and which allows us to identify schools (as workplaces) and teachers. These data will serve two primary purposes; (i) they will allow us to check much longer pre-trends of adopter and non-adopter schools, going back to 2002, in validating the D-in-D design we use, and (ii) they will allow us to explore the role of local labour markets in our analysis, in relation to both the observed pattern of adoption of pay flexibility and in moderating the impacts of the pay reforms on labour market outcomes for teachers.

For this latter work, having merged our school-level indicator of adoption of pay flexibility, derived in the SWF, to workplaces in ASHE, using postcodes available in both sets of data, we will construct bespoke school-specific labour markets based on straight line distances between schools. In ASHE we observe the wages of similarly paid non-teachers, which we will use to construct measures of teacher outside options at the level of the local labour market. In tight labour markets (where outside options are favourable) we anticipate an increase in the real pay of new entrants since teacher wages will adjust to local labour market conditions due to supply and demand for able workers. Conversely, faced with slack labour markets where school wages are above those available among competing employers, we anticipate downward starter wage drift. We will similarly consider the wage progression, exit and entry rates, and the within-occupation mobility of teachers relative to "like" workers in similar professions using ASHE.

References

Biasi, B. (2018) "The Labor Market For Teachers Under Different Pay Schemes", *NBER Working Paper No. 24813*

Britton, J. and Propper, C. (2016): "Teacher pay and school productivity: Exploiting wage regulation," *Journal of Public Economics*, 133(C): 75-89.

Burgess, S., Greaves E. and Murphy R. (2018): "Deregulating teacher labour markets," mimeo

Busby, E. (2015): "Performance-related pay? Not in my school, teachers say" (https://www.tes.com/magazine/article/performance-related-pay-not-my-school-teachers-say)

Chetty, R., Friedman J. and Rockoff, J. (2014a): "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review*, 104 (9): 2593-2632.

Chetty, R., Friedman J. and Rockoff, J. (2014b): "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*, 104 (9): 2633-79.

DfE (2012a): "Evidence to the STRB: the case for change." Department for Education. (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/192218/evidence_to_the_strb_the_case_for_change.pdf)

DfE (2012b): "School Teachers' Pay and Conditions Document 2012 and Guidance on School Teachers' Pay and Conditions." Department for Education (https://webarchive.nationalarchives.gov.uk/20130404004339/https://www.education.gov.uk/publications/eOrderingDownload/School%20Teacher%20Pay%20and%20Conditions%202012.pdf)

DfE (2013): "School Teachers' Pay and Conditions Document 2013 and Guidance on School Teachers' Pay and Conditions." Department for Education (https://media.education.gov.uk/assets/files/pdf/s/130806%202013%20stpcd%20master%20final.pdf)

DfE (2014): "School Teachers' Pay and Conditions Document 2014 and Guidance on School Teachers' Pay and Conditions." Department for Education (https://dera.ioe.ac.uk/20650/1/School_teachers__pay_and_conditions_2014.pdf)

DfE (2018): "The proportion of pupils in academies and free schools, in England, in October 2018" (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/772809/Proportion_of_pupils_in_academies_and_free_schools.pdf)

DfE (2019): "Implementing your school's approach to pay." Department for Education (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/786098/Implementing_your_school_s_approach_to_pay.pdf)

Eyles, A. & Machin, S. (2015) The introduction of academy schools to England's education. IZA Discussion Paper 9276. Bonn: IZA Institute of Labor Economics. http://ftp.iza.org/dp9276.pdf

Loyalka, P., Sylvia, S., Liu, C., Chu, J. and Shi, Y. (forthcoming) "Pay by Design: Teacher Performance Pay Design and the Distribution of Student Achievement", *Journal of Labor Economics*

Manning, A. (2003). *Monopsony in motion: Imperfect competition in labor markets*. Princeton, N.J: Princeton University Press.

Murnane, R. J. and Olsen, R. J. (1989) "The effect of salaries and opportunity costs on duration in teaching: evidence from Michigan", *The Review of Economics and Statistics*, 71, 2: 347-352

NEU (2019): "Performance-related pay for teachers" (https://neu.org.uk/policy/performance-related-pay-teachers)

NUT (2014): "School teachers' pay NUT guidance on the teachers' pay structure for 2014." (http://www.teachers.org.uk/files/nut-guidance-on-the-teachers-pay-structure-2014-15-final.doc)

Propper, C. and Van Reenen, J. (2010) "Can Pay Regulation Kill? Panel data evidence on the effect of labor markets on hospital performance", *Journal of Political Economy*, 118(2): 222-273.

Ransom, M. R. and Sims, D. P. (2010): "Estimating the Firm's Labor Supply Curve in a 'New Monopsony' Framework: Schoolteachers in Missouri." *Journal of Labor Economics,* 28(2): 331-355.

Rockoff, J. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review, 94*(2), 247-252.

Rothstein, J. (2010): “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement,” *The Quarterly Journal of Economics*, Volume 125, Issue 1, 1 February 2010, 175–214

Rothstein, J. (2017): "Measuring the Impacts of Teachers: Comment." *American Economic Review*, 107 (6): 1656-84.

Sharpe et al. (2017): “Evaluation of Teachers’ Pay Reform,” Department for Education (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652763/Evaluation_of_Teachers__Pay_Reform-Final_Report.pdf)

STRB (2013): Twenty-Second Report – 2013 (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/208618/School_Teachers_Review_Body_-_22nd_report.pdf)

Unions (2018): “Joint Advice on School Teachers’ Pay 2018-19” (https://neu.org.uk/sites/default/files/2019-01/2018-19%20Pay%20Scale%20Joint%20Advice.pdf)

Willen, A. (2019) “Decentralization of Wage Determination: Evidence from a National Teacher Reform”, mimeo

Figures

Figure 1: Example for binding, STRB-designated pay schedules

Panel A: Pre-reform

Pay Scale for Classroom Teachers (Main scale) 2012				
Scale point	*Annual Salary England and Wales (excluding the London Area)*	*Annual Salary Inner London Area*	*Annual Salary Outer London Area*	*Annual Salary Fringe Area*
	£	£	£	£
1	21,588	27,000	25,117	22,626
2	23,295	28,408	26,674	24,331
3	25,168	29,889	28,325	26,203
4	27,104	31,446	30,080	28,146
5	29,240	33,865	32,630	30,278
6	31,552	36,387	35,116	32,588

Pay Scale for Post-Threshold Teachers 2012				
Scale point	*Annual Salary England and Wales (excluding the London Area)*	*Annual Salary Inner London Area*	*Annual Salary Outer London Area*	*Annual Salary Fringe Area*
	£	£	£	£
U1	34,181	41,497	37,599	35,218
U2	35,447	43,536	38,991	36,483
U3	36,756	45,000	40,433	37,795

Panel B: Post-reform

Main Pay Range 2014				
	Annual Salary			
	England and Wales (excluding the London Area) £	*Inner London Area* £	*Outer London Area* £	*Fringe Area* £
Minimum	22,023	27,543	25,623	23,082
Maximum	32,187	37,119	35,823	33,244

Upper Pay Range 2014				
	Annual Salary			
	England and Wales (excluding the London Area) £	*Inner London Area* £	*Outer London Area* £	*Fringe Area* £
Minimum	34,869	42,332	38,355	35,927
Maximum	37,496	45,905	41,247	38,555

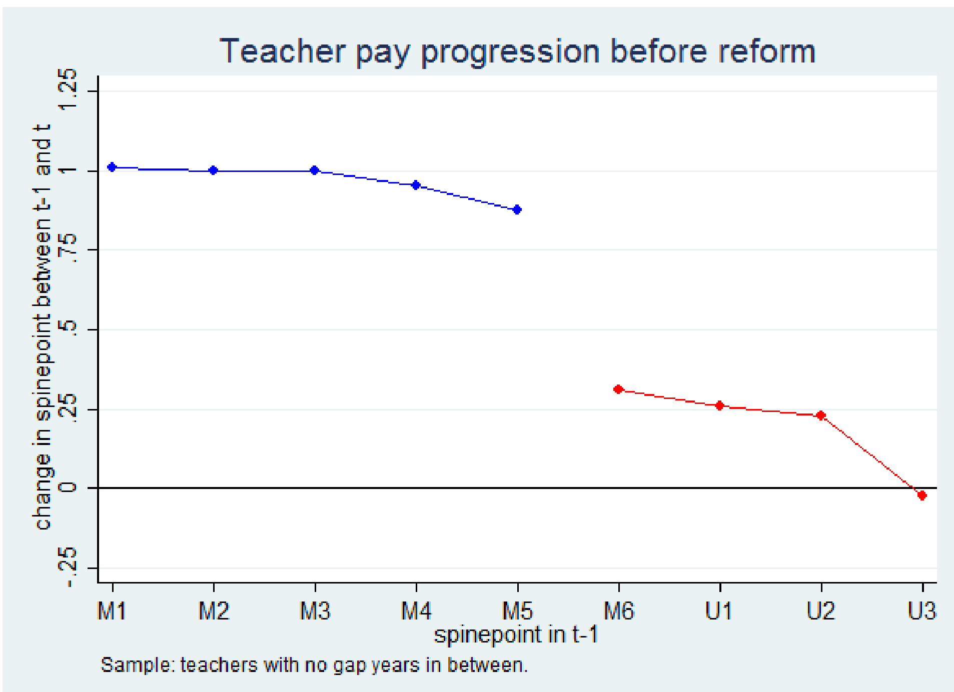
Source: pp. 56,60 in DfE (2012b); pp. 19-20 in DfE (2014)

Figure 2: Example for non-binding, union-designated pay schedules, post-reform

	England & Wales 1 Sept 2014	Inner London 1 Sept 2014	Outer London Area 1 Sept 2014	Fringe 1 Sept 2014
Main Pay Range	**£ p.a.**	**£ p.a.**	**£ p.a.**	**£ p.a.**
Minimum *M1*	22,023	27,543	25,623	23,082
M2	*23,764*	*28,980*	*27,211*	*24,821*
M3	*25,675*	*30,490*	*28,896*	*26,731*
M4	*27,650*	*32,079*	*30,685*	*28,713*
M5	*29,829*	*34,547*	*33,287*	*30,887*
Maximum *M6*	32,187	37,119	35,823	33,244
Upper Pay Range	**£ p.a.**	**£ p.a.**	**£ p.a.**	**£ p.a.**
Minimum	34,869	42,332	38,355	35,927
Maximum	37,496	45,905	41,247	38,555

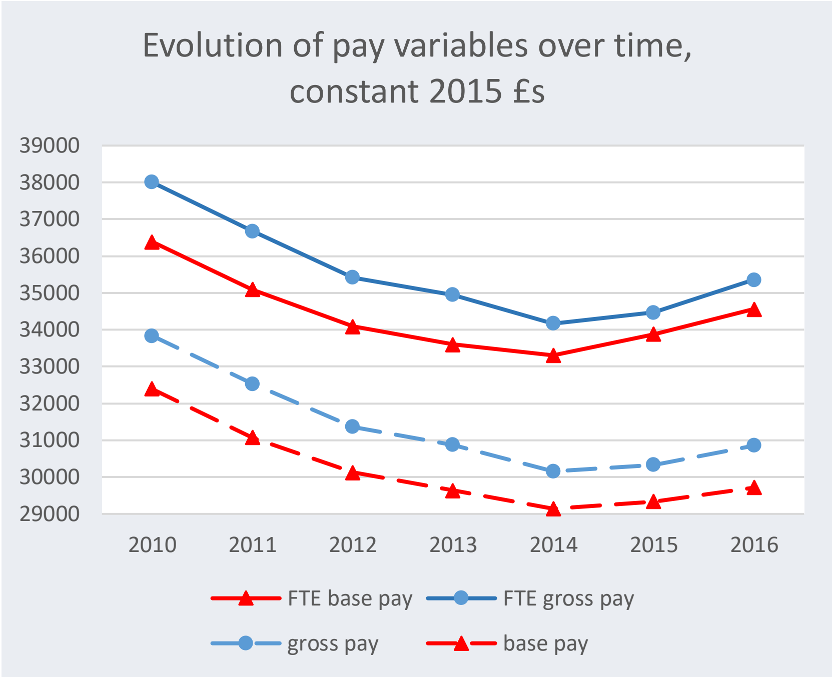
Source: p. 1 in NUT 2014

Figure 3: Progression of teachers through the main and upper pay scales, 2010-13



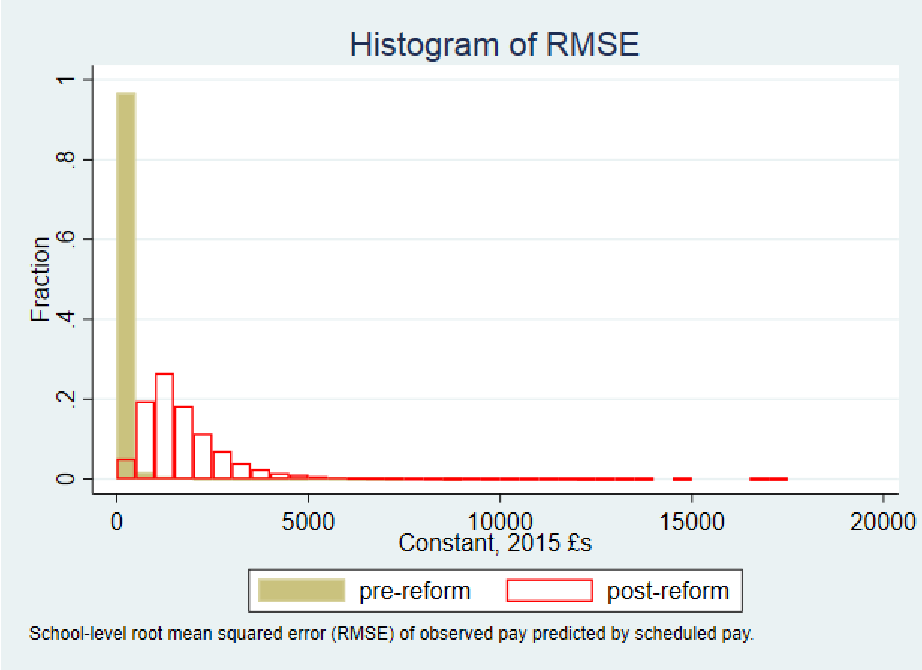
Note: Average progression at each spine point pre-reform. The chart shows that teachers at the bottom and middle of the pay ladder (relatively young teachers with little experience) automatically progress to the next spine point in the following year. More experienced teachers progress approximately once every four year.

Figure 4: Evolution of pay variables over time



Note: Evolution of the mean of key pay variables in the SWF dataset. Full-time equivalent base pay is centrally regulated, while gross pay includes additional payments received for taking up extra responsibilities.

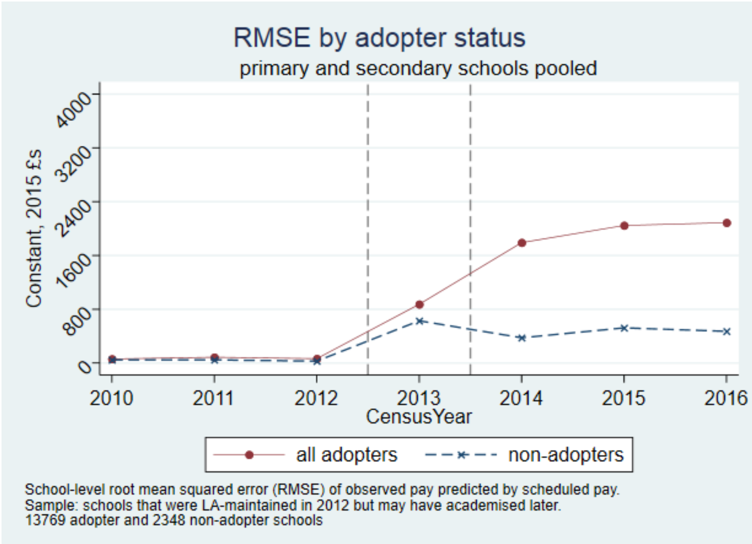
Figure 5: Histogram of RMSE pre- and post-reform



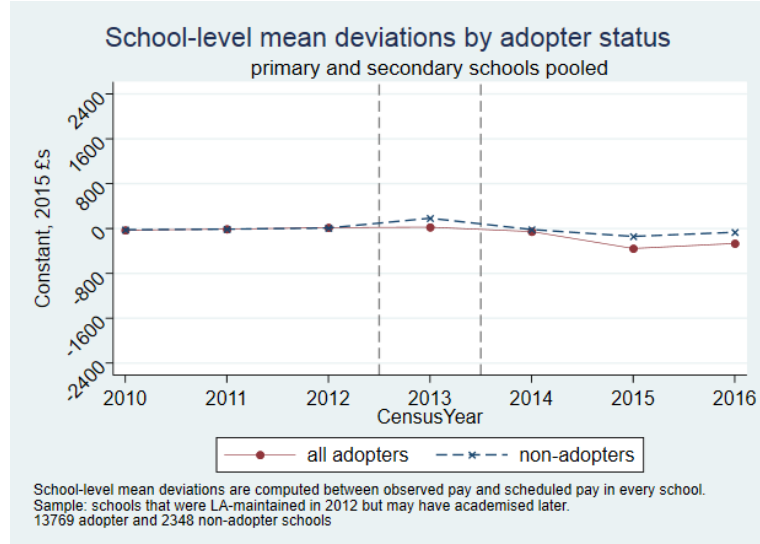
Note: School-level root mean squared error (RMSE) of observed full-time equivalent base pay when predicted by scheduled pay (what would be expected based on the pre-reform seniority pay regime).

Figure 6: Classification diagnostics – Evolution of RMSE and mean deviation over time by adopter status

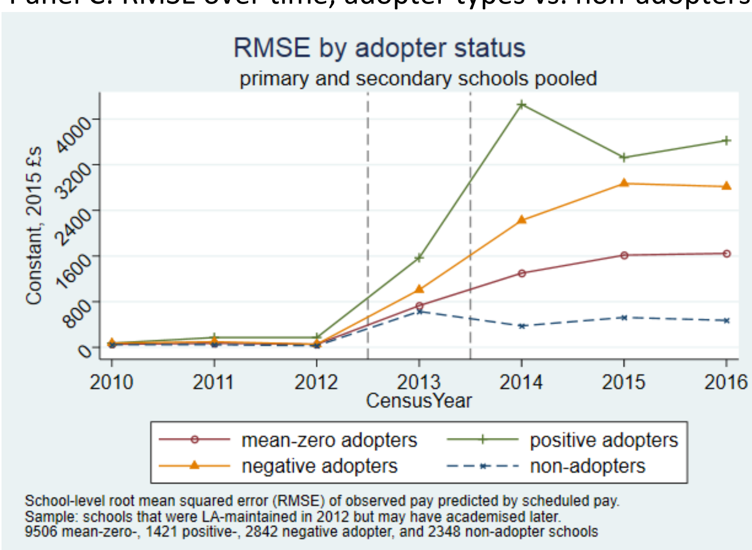
Panel A: RMSE over time, adopters vs. non-adopters



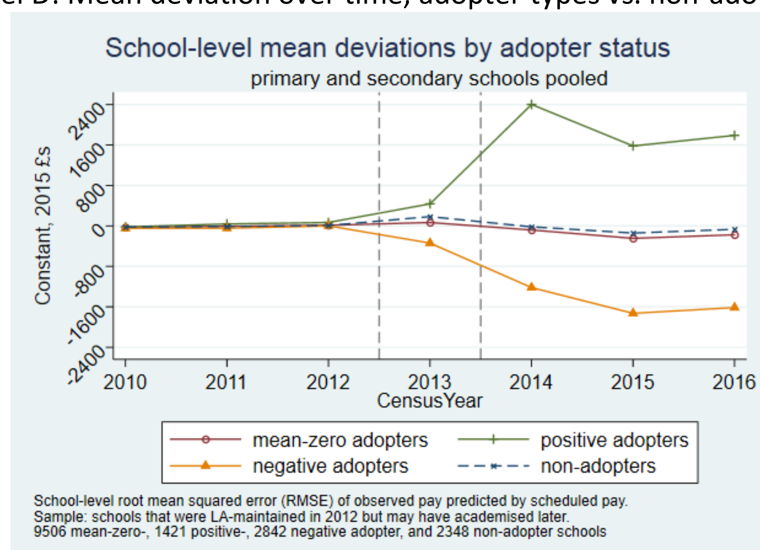
Panel B: Mean deviation over time, adopters vs. non-adopters



Panel C: RMSE over time, adopter types vs. non-adopters



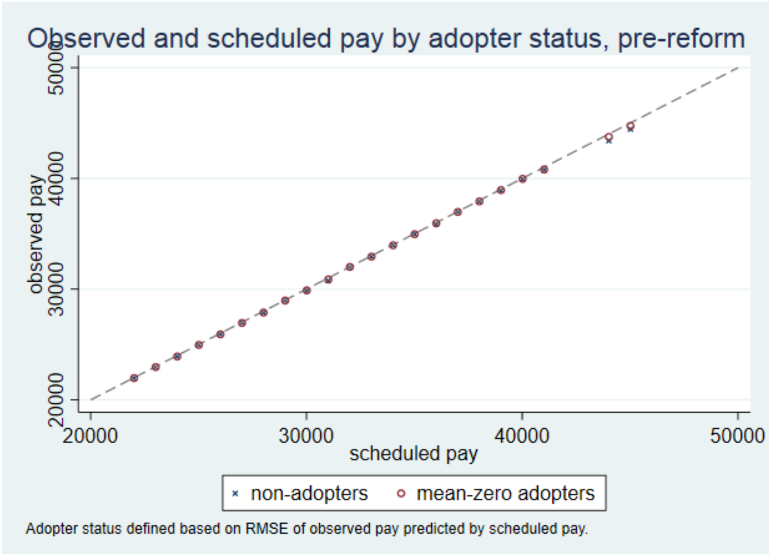
Panel D: Mean deviation over time, adopter types vs. non-adopters



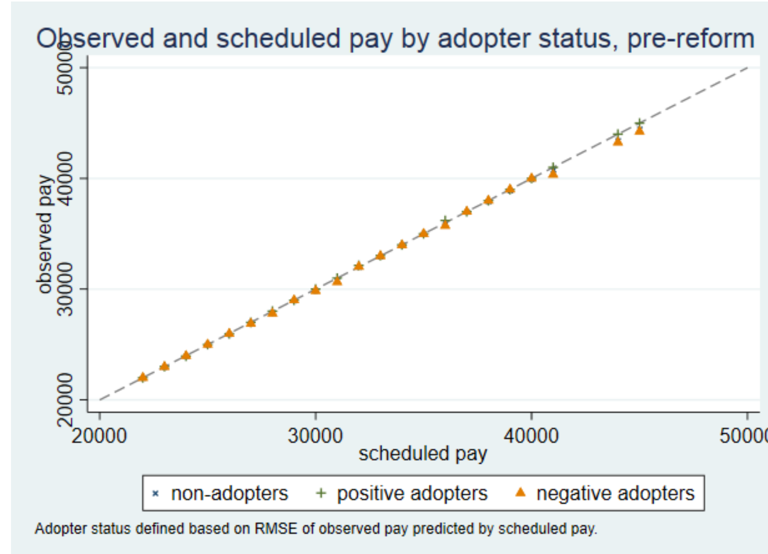
Note: Each panel shows differences between schools of different adopter status in terms of (i) RMSE of observed pay when predicted by scheduled pay (what would be expected based on pre-reform seniority pay) and (ii) school-level mean deviation between observed and scheduled pay. These differences in the post-reform period are by construction, as we use them to define adopter status. See more details in Section 4.

Figure 7: Empirical pay schedule by adopter status, pre- and post-reform

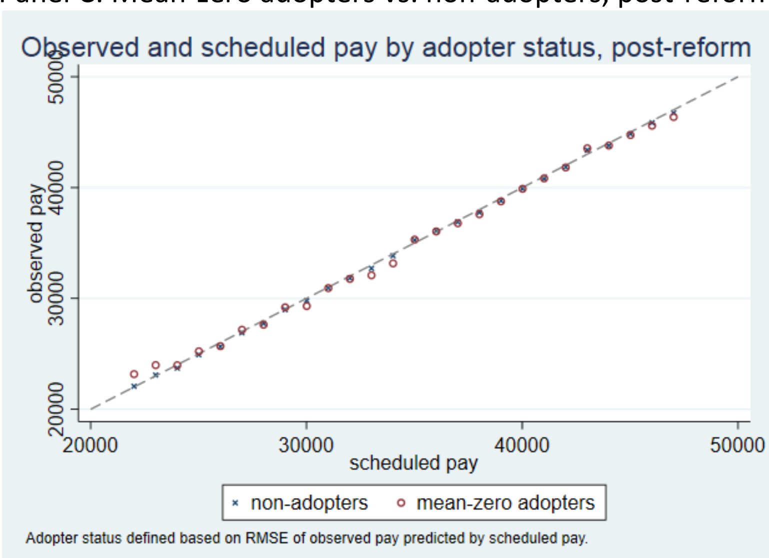
Panel A: Mean-zero adopters vs. non-adopters, pre-reform



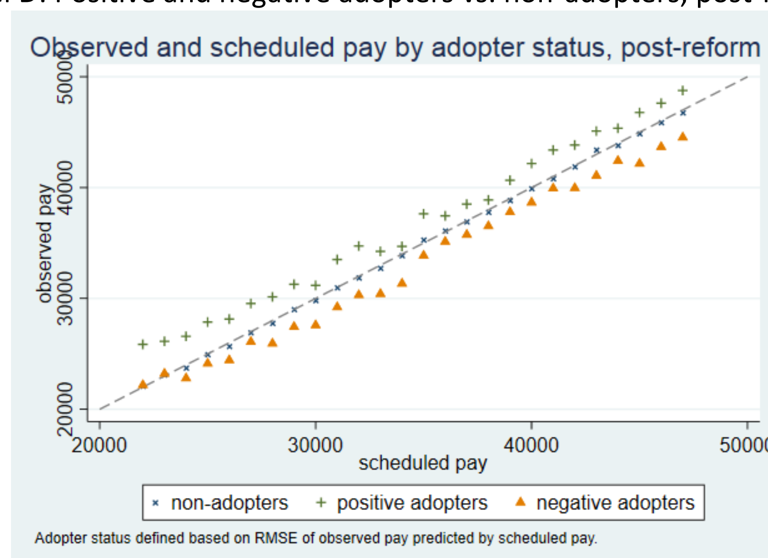
Panel B: Positive and negative adopters vs. non-adopters, pre-reform



Panel C: Mean-zero adopters vs. non-adopters, post-reform



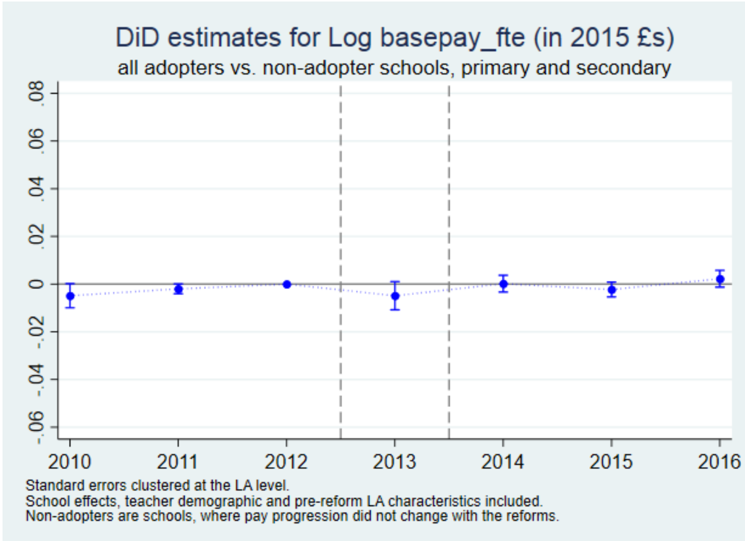
Panel D: Positive and negative adopters vs. non-adopters, post-reform



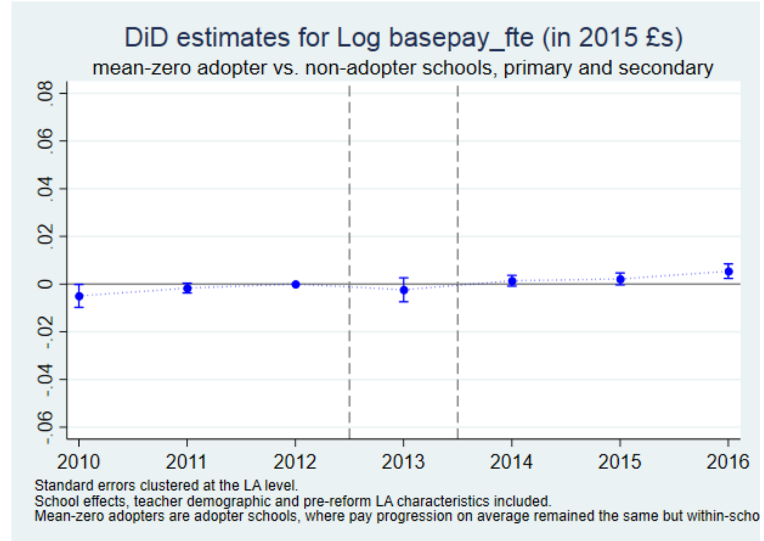
Note: Each panel shows a nonparametric relationship between scheduled pay (what would be expected under pre-reform seniority pay) and observed pay. The panels illustrate how each adopter type departs/does not depart from scheduled pay pre- and post-reform.

Figure 8: The effect of the pay reform on teacher pay – difference-in-difference estimates for full-time equivalent base pay

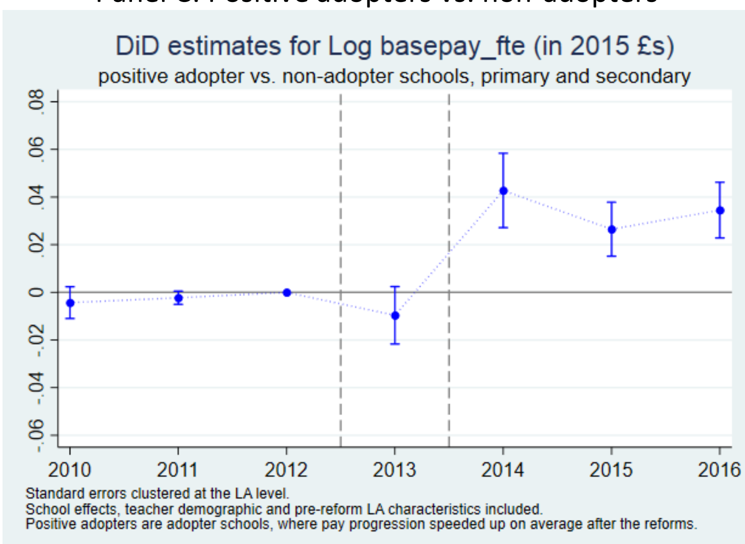
Panel A: All adopters vs. non-adopters



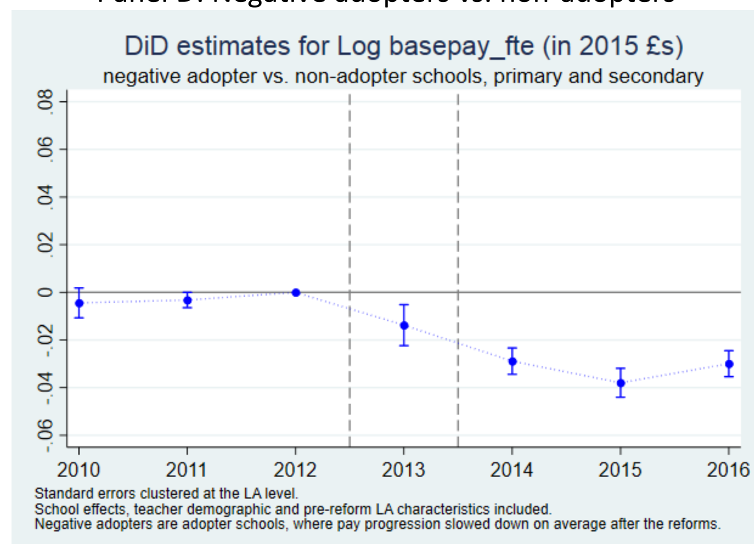
Panel B: Mean-zero adopters vs. non-adopters



Panel C: Positive adopters vs. non-adopters



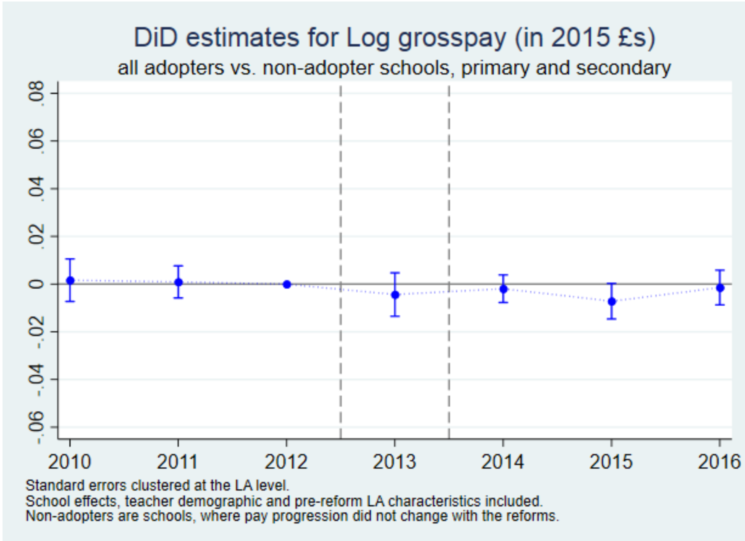
Panel D: Negative adopters vs. non-adopters



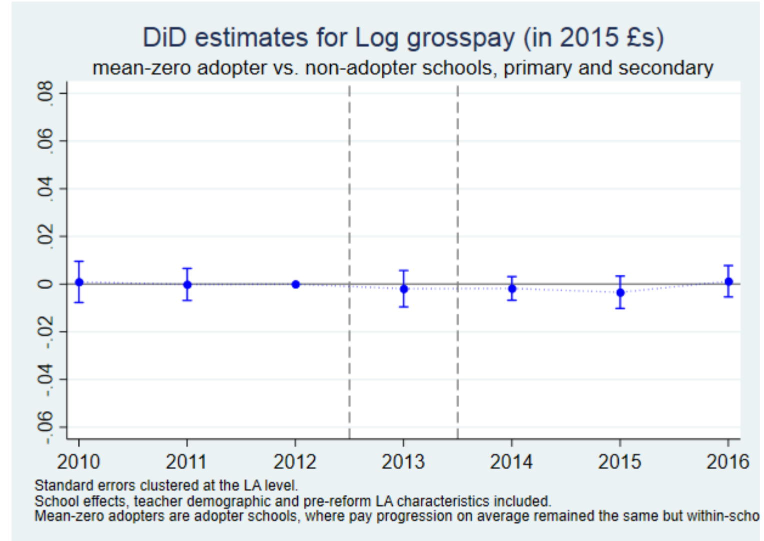
Note: Each panel shows the estimates from difference-in-difference model in Section 5.1 for teachers in the given type of adopter schools against those in non-adopter schools. The dependent variable is full-time equivalent base pay. Standard errors are clustered at the LA level, and all models include school effects and teacher demographic characteristics and pre-reform pay points. See more details about adopter types in Section 4.

Figure 9: The effect of the pay reform on teacher pay – difference-in-difference estimates for gross pay

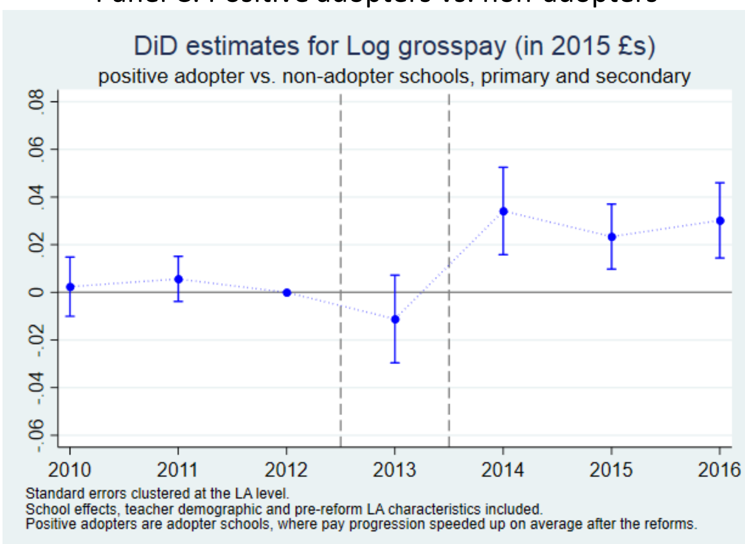
Panel A: All adopters vs. non-adopters



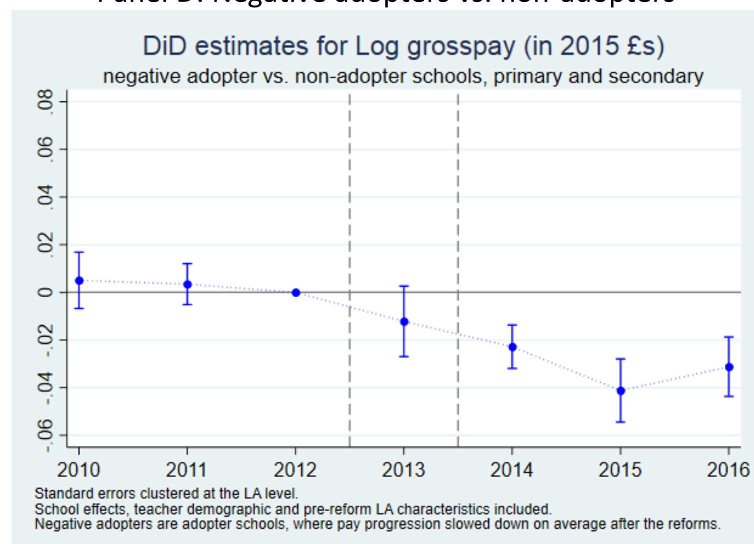
Panel B: Mean-zero adopters vs. non-adopters



Panel C: Positive adopters vs. non-adopters



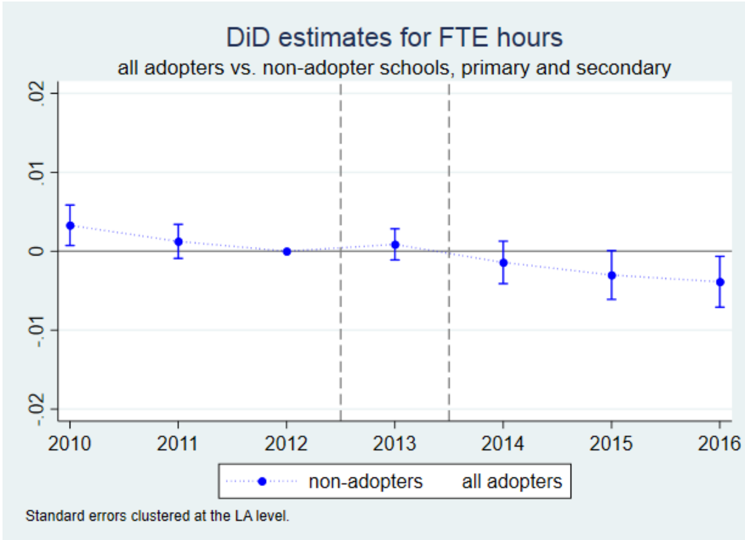
Panel D: Negative adopters vs. non-adopters



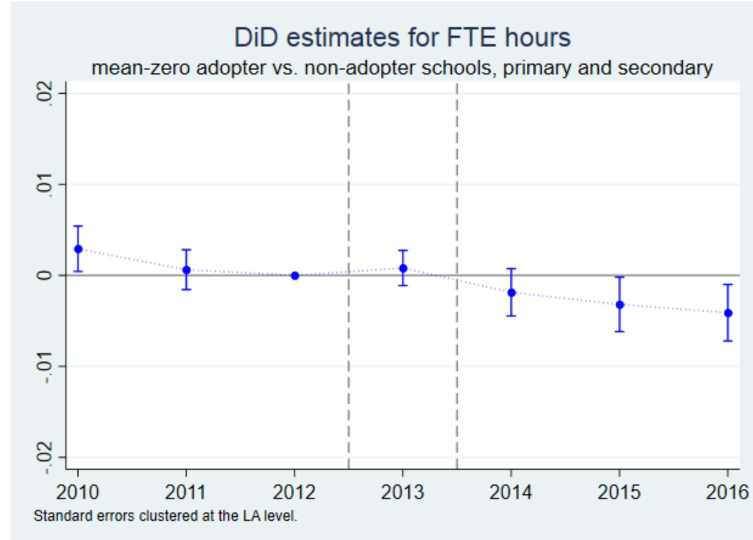
Note: Each panel shows the estimates from difference-in-difference model in Section 5.1 for teachers in the given type of adopter schools against those in non-adopter schools. The dependent variable is gross pay. Standard errors are clustered at the LA level, and all models include school effects and teacher demographic characteristics and pre-reform pay points. See more details about adopter types in Section 4.

Figure 10: The effect of the pay reform on teacher pay – difference-in-difference estimates for full-time equivalent hours

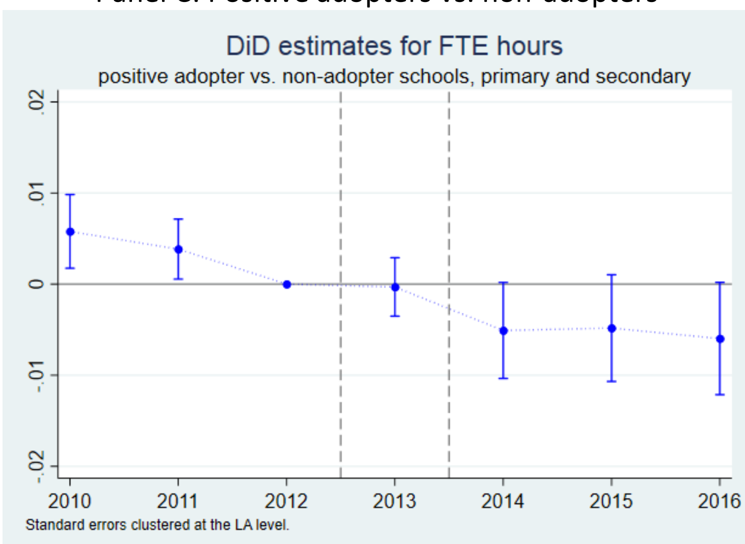
Panel A: All adopters vs. non-adopters



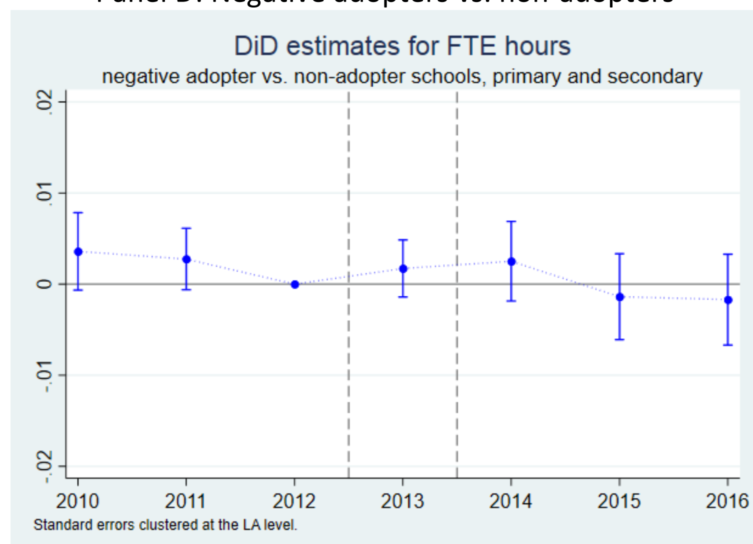
Panel B: Mean-zero adopters vs. non-adopters



Panel C: Positive adopters vs. non-adopters



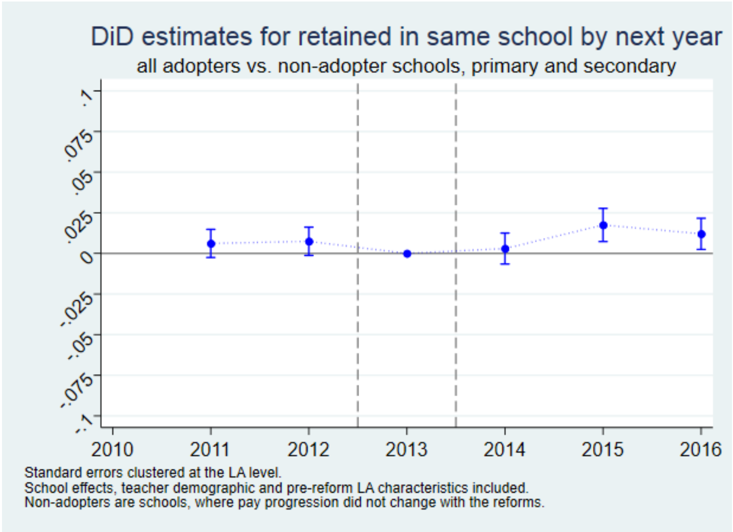
Panel D: Negative adopters vs. non-adopters



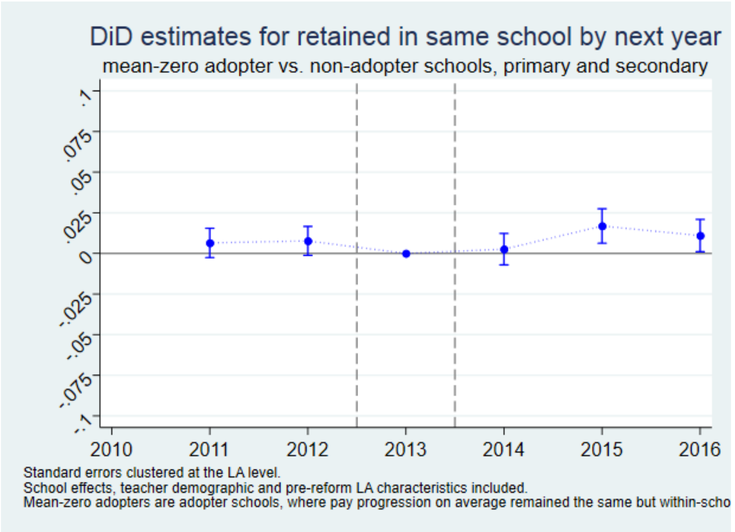
Note: Each panel shows the estimates from difference-in-difference model in Section 5.1 for teachers in the given type of adopter schools against those in non-adopter schools. The dependent variable is full-time equivalent hours. Standard errors are clustered at the LA level, and all models include school effects and teacher demographic characteristics and pre-reform pay points. See more details about adopter types in Section 4.

Figure 11: The effect of the pay reform on teacher pay – difference-in-difference estimates for teacher retention

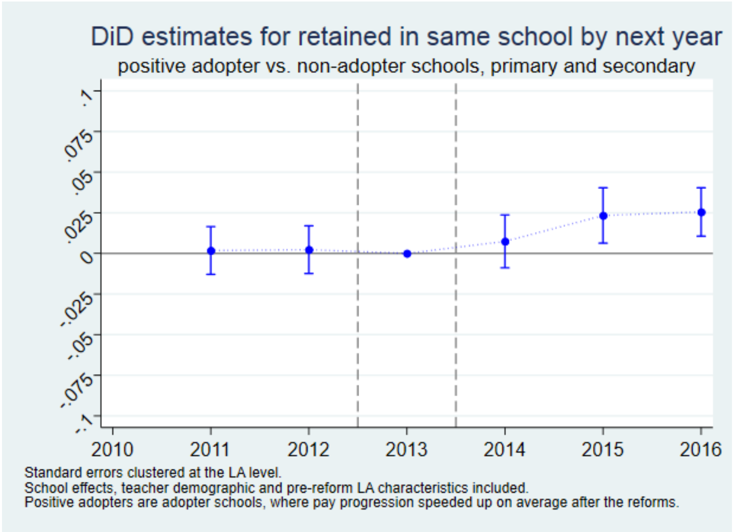
Panel A: All adopters vs. non-adopters



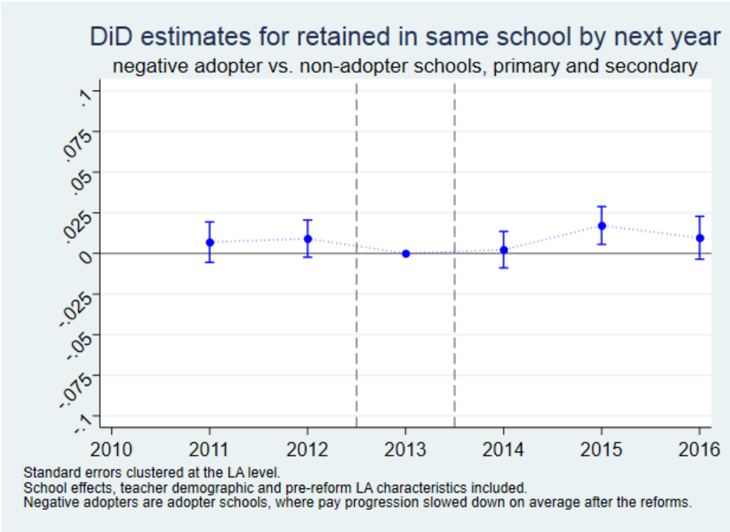
Panel B: Mean-zero adopters vs. non-adopters



Panel C: Positive adopters vs. non-adopters



Panel D: Negative adopters vs. non-adopters



Note: Each panel shows the estimates from difference-in-difference model in Section 5.1 or teachers in the given type of adopter schools against those in non-adopter schools. The dependent variable is an indicator if the teacher stayed in the same school for the given year. Standard errors are clustered at the LA level, and all models include school effects and teacher demographic characteristics and pre-reform pay points. See more details about adopter types in Section 4.

Tables

Table 1: Summary statistics – Teachers

	mean	sd	min	max	N
male	0.21				453,518
white British	0.81				454,020
age	35.75	11.14	16	88	454,005
under 30	0.42				454,005
over 45	0.24				454,005
M1	0.10				454,020
M2	0.05				454,020
M3	0.04				454,020
M4	0.04				454,020
M5	0.03				454,020
M6	0.10				454,020
U1	0.08				454,020
U2	0.07				454,020
U3	0.15				454,020
other spine point	0.07				454,020
missing spine point	0.01				454,020
not in sample in 2012	0.26				454,020
number of years in sample	4.00	2.23	1	7	454,020
avg FTE	0.89	0.20	0.00	1.20	454,020
number of schools worked at	1.21	0.47	1	6	454,020

Note: Summary statistics from teachers working in base sample schools in any given year (School Workforce Census). Age and spine point statistics reported are from the first year when teacher appears in sample.

Table 2: Summary statistics – Schools

	mean	sd	min	max	N
years in data	6.96	0.25	4.00	7.00	16,130
secondary	0.09				16,130
foundation	0.07				16,130
vol aided	0.23				16,130
vol controlled	0.14				16,130
academise after 2012	0.17				16,130
academy converter after 2012	0.12				16,130
sponsor-led academy after 2012	0.05				16,130
fringe London	0.06				16,130
inner London	0.06				16,130
outer London	0.05				16,130
urban	0.71				16,130
FTE teachers	13.89	15.97	0.20	158.46	16,130
FTE pupils	306.12	261.83	8	2191	15,852
teacher turnover	0.18	0.16	0.00	1.00	16,130
%FSM	17.89	13.76	0.00	75.00	15,280
exp on teachers	2162.83	454.45	1080.16	5767.92	14,927
total exp per pupil	4444.21	1092.98	2594.70	40023.73	15,853
exp ratio on teaching	0.65	0.16	0.00	1.06	15,853
budget balance	42.17	211.42	-9881.96	2496.68	15,853
ratio of grant income	0.97	0.03	0.41	1.05	15,853
M1 share	0.08	0.10	0.00	1.00	16,130
M2 share	0.06	0.09	0.00	1.00	16,130
M3 share	0.06	0.08	0.00	1.00	16,130
M4 share	0.05	0.08	0.00	1.00	16,130
M5 share	0.05	0.08	0.00	1.00	16,130
M6 share	0.16	0.15	0.00	1.00	16,130
U1 share	0.15	0.14	0.00	1.00	16,130
U2 share	0.11	0.13	0.00	1.00	16,130
U3 share	0.22	0.19	0.00	1.00	16,130
other spine point share	0.04	0.10	0.00	0.97	16,130
missing spine point share	0.01	0.06	0.00	0.88	16,130

Note: 2012 summary statistics of base sample schools (School Workforce Census and Consistent Finance Returns)

Table 3: Number of schools by adopter status

	Primary	Secondary	All
Non-adopter	2,225	123	2,348
Mean-zero adopter	8,402	1,104	9,506
Positive adopter	1,317	104	1,421
Negative adopter	2,669	173	2,842
Total	14,613	1,504	16,117

Note: Table shows the adopter classification resulting from the procedure described in detail in Section 4.1. Non-adopters are schools where pay progression did not change after the reforms. Mean-zero adopters are schools where on average, pay progression did not change after the reforms but still, some teachers progressed faster, while some slower. Positive adopters are schools where pay progression speeded up on average after the reforms, while in negative adopter schools, it slowed down on average.

Table 4: Predictors of adopting some form of flexibility

VARIABLES	(1) all adopters	(2) mean-zero adopters	(3) positive adopters	(4) negative adopters
secondary	0.092**		0.189*	3.432***
	(0.046)		(0.113)	(1.286)
urban	0.024***	0.028**		
	(0.008)	(0.011)		
2010-12 avg log fte pupils	0.071***	0.141***		
	(0.013)	(0.015)		
2010-12 avg growth of fte pupils	0.116*			0.375**
	(0.059)			(0.148)
2010-12 avg growth of % FSM	-0.020		-0.049	-0.069*
	(0.017)		(0.034)	(0.037)
2010-12 avg KS lagged score	-0.005*		-0.016*	-0.011**
	(0.003)		(0.009)	(0.005)
2010-12 avg log exp on teachers	0.034	0.062*		
	(0.035)	(0.033)		
2010-12 avg log exp on supply teachers	0.009**	0.013**		0.020*
	(0.004)	(0.006)		(0.011)
2010-12 avg growth of grant income	-0.139*	-0.275***		
	(0.074)	(0.095)		
2010-12 avg teacher FTE	-0.001*	-0.002***		0.003**
	(0.000)	(0.001)		(0.001)
2010-12 avg teacher turnover	0.153***	0.211***	0.130	0.348***
	(0.036)	(0.046)	(0.095)	(0.087)
2010-12 avg male teacher share	-0.038	-0.019		-0.124*
	(0.028)	(0.037)		(0.072)
2010-12 avg full-time teacher share	0.051**	0.042		0.113***
	(0.020)	(0.026)		(0.042)
urn201012_age2010-12 avg teacher age	0.013***	0.015***	0.012**	0.007***
	(0.003)	(0.003)	(0.005)	(0.002)
2010-12 avg below-30 teacher share	0.120**	0.147***		
	(0.047)	(0.048)		
2010-12 avg above-45 teacher share	-0.140**	-0.174***	-0.088	
	(0.056)	(0.058)	(0.091)	
2010-12 avg M3 share	-0.213***	-0.284***	-0.816***	-0.585***
	(0.064)	(0.097)	(0.199)	(0.181)
2010-12 avg M4 share	-0.289***	-0.396***	-1.438***	-0.736***
	(0.072)	(0.095)	(0.225)	(0.163)
2010-12 avg M5 share	-0.171***	-0.255***	-1.133***	-0.785***
	(0.061)	(0.093)	(0.200)	(0.194)
2010-12 avg M6 share	-0.245***	-0.303***	-1.192***	-0.694***
	(0.041)	(0.057)	(0.165)	(0.142)

2010-12 avg U1 share	-0.197***	-0.246***	-1.029***	-0.721***
	(0.041)	(0.057)	(0.174)	(0.145)
2010-12 avg U2 share	-0.380***	-0.473***	-1.320***	-1.008***
	(0.048)	(0.065)	(0.175)	(0.143)
2010-12 avg U3 share	-0.438***	-0.543***	-1.366***	-1.071***
	(0.044)	(0.060)	(0.161)	(0.139)
2010-12 avg other paypoint share	-0.417***	-0.570***	-1.177***	-0.917***
	(0.049)	(0.065)	(0.145)	(0.127)
2010-12 avg missing paypoint share	-0.239***	-0.390***	-0.688***	-0.886***
	(0.057)	(0.079)	(0.212)	(0.215)
female mean %change in annual pay in LA (2012)	-0.005***	0.005***	-0.010**	0.019***
	(0.001)	(0.001)	(0.004)	(0.004)
log avg annual pay in LA (2012)	0.077***	-0.018	-0.180***	0.236***
	(0.024)	(0.023)	(0.064)	(0.031)
mean %change in annual pay in LA (2012)	0.006***	-0.002***	-0.015***	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Inner London		0.025	-0.115***	
		(0.025)	(0.042)	
Outer London		0.123***	-0.880***	
		(0.013)	(0.105)	
2010-12 avg % FSM		0.001*	0.001	
		(0.000)	(0.001)	
share of academies in LA		0.011		
		(0.034)		
vol. aided			-0.027	
			(0.017)	
Fringe London			-0.101	
			(0.098)	
2010-12 avg KS score			0.003	
			(0.003)	
2010-12 avg log grant income			-0.091*	
			(0.049)	
2010-12 avg white British teacher share			-0.241***	
			(0.086)	
2010-12 avg M2 share			-0.643***	-0.343
			(0.217)	(0.212)
2010-12 avg KS VA				-0.004**
				(0.001)
2010-12 avg % budget balance				-0.411
				(0.324)
Constant	-0.712***	-0.660**	4.229***	-1.019***
	(0.157)	(0.324)	(0.818)	(0.348)
Observations	13,023	11,347	2,835	3,948
R-squared	0.114	0.128	0.544	0.303

phase	primary and secondary	primary and secondary	primary and secondary	primary and secondary
LA effects	yes	yes	yes	yes
joint p-value for LA effects	0.000	0.000	0.000	0.000

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

Standard errors, clustered at the LA level, in parentheses.

OLS regression model selected based on a backward AIC criterion. Initial set of predictors: general school characteristics (school type, region, urban dummy); 2010-12 school-level average of pupil- and teacher characteristics (log pupil number and its growth rate, % FSM and its growth rate, KS scores, VA and lagged scores, teacher number and its growth rate, turnover and its growth rate, teacher shares of white British, males, below 30, above 45); 2010-12 average of school budget items (real expenditure on teachers, supplement teachers, education support staff, real self-generated/grant income; difference of log real income and expenditures); 2012 LA-level average of annual pay, its % change for everyone and for females only; share of academies in LA.

Table 5: Difference-in-difference estimates for full-time equivalent base pay

VARIABLES	(1) all adopters	(2) mean-zero adopters	(3) positive adopters	(4) negative adopters
2010	0.079***	0.079***	0.078***	0.079***
	(0.003)	(0.003)	(0.002)	(0.003)
2011	0.030***	0.030***	0.029***	0.030***
	(0.001)	(0.001)	(0.001)	(0.001)
2013	0.020***	0.020***	0.020***	0.020***
	(0.005)	(0.005)	(0.005)	(0.005)
2014	0.021***	0.021***	0.021***	0.020***
	(0.002)	(0.002)	(0.002)	(0.002)
2015	0.049***	0.049***	0.048***	0.048***
	(0.002)	(0.002)	(0.002)	(0.002)
2016	0.060***	0.060***	0.059***	0.058***
	(0.002)	(0.002)	(0.002)	(0.002)
2010*adopter	-0.005*	-0.005**	-0.004	-0.004
	(0.003)	(0.002)	(0.003)	(0.003)
2011*adopter	-0.002*	-0.002	-0.002	-0.003*
	-0.001	(0.001)	(0.001)	(0.002)
2013*adopter	-0.005	-0.002	-0.010	-0.014***
	(0.003)	(0.003)	(0.006)	(0.004)
2014*adopter	0.000	0.001	0.043***	-0.029***
	(0.002)	-0.001	(0.008)	(0.003)
2015*adopter	-0.002	0.002*	0.027***	-0.038***
	-0.002	(0.001)	(0.006)	-0.003
2016*adopter	0.002	0.005***	0.035***	-0.030***
	(0.002)	(0.002)	(0.006)	(0.003)
Observations	1,713,844	1,325,204	325,193	440,127
R-squared	0.719	0.730	0.702	0.709
teacher chars	yes	yes	yes	yes
school effects	yes	yes	yes	yes

Standard errors, clustered at the LA level, in parentheses.

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

Dependent variable is full-time equivalent base pay (at constant, 2015 £s), which was the pay variable subject to the seniority schedule pre-reform. See definitions for the schools of each adopter type in Section 4.

Table 6: Difference-in-difference estimates for full-time equivalent gross pay

VARIABLES	(1) all adopters	(2) mean-zero adopters	(3) positive adopters	(4) negative adopters
2010	0.081***	0.082***	0.080***	0.081***
	(0.005)	(0.005)	(0.005)	(0.005)
2011	0.033***	0.033***	0.033***	0.033***
	(0.003)	(0.003)	(0.003)	(0.003)
2013	0.026***	0.026***	0.026***	0.025***
	(0.009)	(0.009)	(0.009)	(0.009)
2014	0.031***	0.032***	0.031***	0.028***
	(0.003)	(0.003)	(0.004)	(0.004)
2015	0.064***	0.065***	0.063***	0.060***
	(0.005)	(0.005)	(0.005)	(0.005)
2016	0.076***	0.077***	0.075***	0.072***
	(0.005)	(0.005)	(0.006)	(0.005)
2010*adopter	0.002	0.001	0.002	0.005
	(0.005)	(0.004)	(0.006)	(0.006)
2011*adopter	0.001	-0.000	0.006	0.003
	(0.003)	(0.003)	(0.005)	(0.004)
2013*adopter	-0.004	-0.002	-0.011	-0.012
	(0.005)	(0.004)	-0.009	(0.008)
2014*adopter	-0.002	-0.002	0.034***	-0.023***
	(0.003)	(0.003)	(0.009)	(0.005)
2015*adopter	-0.007*	-0.003	0.023***	-0.041***
	(0.004)	(0.003)	(0.007)	(0.007)
2016*adopter	-0.001	0.001	0.030***	-0.031***
	(0.004)	(0.003)	(0.008)	(0.006)
Observations	1,713,844	1,325,204	325,193	440,127
R-squared	0.250	0.247	0.249	0.244
teacher chars	yes	yes	yes	yes
school effects	yes	yes	yes	yes

Standard errors, clustered at the LA level, in parentheses.

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

Dependent variable is gross pay (at constant, 2015 £s), which is the sum of FTE base pay (subject to seniority schedule pre-reform) multiplied by full-time equivalent hours and additional pay elements received for taking up extra responsibilities. See definitions for the schools of each adopter type in Section 4.

Table 7: Difference-in-difference estimates for full-time equivalent hours

VARIABLES	(1) all adopters	(2) mean-zero adopters	(3) positive adopters	(4) negative adopters
2010	-0.000	0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
2011	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
2013	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
2014	0.000	0.000	0.000	-0.000
	(0.001)	(0.001)	(0.002)	(0.001)
2015	-0.001	-0.000	-0.000	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
2016	-0.001	-0.001	-0.000	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
2010*adopter	0.003**	0.003**	0.006***	0.004*
	(0.001)	(0.001)	(0.002)	(0.002)
2011*adopter	0.001	0.001	0.004**	0.003
	(0.001)	(0.001)	-0.002	(0.002)
2013*adopter	0.001	0.001	-0.000	0.002
	(0.001)	-0.001	(0.002)	(0.002)
2014*adopter	-0.001	-0.002	-0.005*	0.003
	(0.001)	(0.001)	(0.003)	(0.002)
2015*adopter	-0.003*	-0.003**	-0.005	-0.001
	(0.002)	(0.002)	(0.003)	-0.002
2016*adopter	-0.004**	-0.004**	-0.006*	-0.002
	-0.002	(0.002)	-0.003	(0.003)
Observations	1,716,979	1,327,172	325,908	441081
R-squared	0.197	0.196	0.214	0.212
teacher chars	yes	yes	yes	yes
school effects	yes	yes	yes	yes

Standard errors, clustered at the LA level, in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Dependent variable: indicator for having been retained from previous year. See definitions for the schools of each adopter type in Section 4.

Table 8: Difference-in-difference estimates for teacher retention

VARIABLES	(1) all adopters	(2) mean-zero adopters	(3) positive adopters	(4) negative adopters
2011	0.025***	0.025***	0.024***	0.025***
	(0.005)	(0.005)	(0.005)	-0.005
2012	0.013***	0.013***	0.013***	0.013***
	(0.004)	(0.004)	(0.004)	-0.004
2014	-0.008*	-0.007*	-0.008*	-0.008*
	(0.004)	(0.004)	(0.004)	-0.004
2015	-0.022***	-0.022***	-0.023***	-0.023***
	(0.005)	(0.005)	(0.005)	-0.005
2016	-0.007	-0.006	-0.007*	-0.008*
	(0.004)	(0.004)	(0.004)	-0.004
2011*adopter	0.006	0.006	0.002	0.007
	(0.004)	(0.005)	(0.007)	-0.006
2012*adopter	0.007*	0.008*	0.002	0.009
	(0.004)	-0.005	(0.007)	(0.006)
2014*adopter	0.003	0.003	0.007	0.002
	(0.005)	(0.005)	(0.008)	-0.006
2015*adopter	0.017***	0.017***	0.023***	0.017***
	(0.005)	(0.005)	-0.009	-0.006
2016*adopter	0.012**	0.011**	0.025***	0.01
	(0.005)	(0.005)	(0.008)	-0.007
Observations	1,469,967	1,136,834	279,304	378,071
R-squared	0.063	0.063	0.071	0.069
teacher chars	yes	yes	yes	yes
school effects	yes	yes	yes	yes

Standard errors, clustered at the LA level, in parentheses.

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

Dependent variable: indicator for having been retained from previous year. See definitions for the schools of each adopter type in Section 4.