Public Sector Performance Disclosure: Salary and Career Outcomes for Top Managers and Employees*

Iftikhar Hussain† Vincenzo Scrutinio‡ Shqiponja Telhaj§

25th September 2023

Abstract

Schools around the world are held to account on the basis of objective measures of performance. This paper investigates the impact of *subjective* school inspection grades on labor market and career outcomes for school principals, senior managers and teachers. We compare personnel in schools experiencing a grade change with those in schools with a stable evaluation, in a difference-in-differences framework. We find that a change in the overall school inspection rating has substantial impact on principals' wages. Investigating other margins of response, such as demotion, turnover, the rate of exit from public sector schooling, and early retirement, reveals even more dramatic effects. Importantly, we also exploit direct inspection ratings on leadership and management quality in order to assess the impact on principals arising from this direct signal, over and above the response to overall school rating. Our findings demonstrate that other senior managers and teachers are also impacted by inspection grade changes, although these effects are much more muted, suggesting that the principal 'carries the can,' with lower ranked managers and teachers left largely unscathed. These results shed new light on the impact of subjective quality assessment on the careers of public sector managers and employees.

J.E.L. codes: I20, I28, J01, J30

Keywords: School Principals; Teacher Labor Market; Quality Disclosure; Inspections.

*Acknowledgments: We would like to thank Department for Education for providing access to the teacher administrative data, especially the Data for Teacher Records team for their invaluable and continuous support since the start of this project. We also thank participants in a number of seminar and conference presentations for comments on earlier drafts of the paper.

†University of Sussex; e-mail: iftikhar.hussain@sussex.ac.uk

‡University of Bologna, IZA & Centre for Economic Performance, London School of Economics and Political Science; e-mail: vincenzo.scrutinio@unibo.it

§Corresponding Author, Department of Economics, University of Sussex & Centre for Economic Performance, London School of Economics and Political Science, e-mail: S.Telhaj@sussex.ac.uk/ S.Telhaj@lse.ac.uk

1 Introduction

Governments around the world make extensive use of performance indicators in the public sector. These are viewed as an important tool for increasing quality in a sector where there is a lack of direct competition and limited accountability (Dranove and Jin, 2010). Publishing performance indicators of public sector organizations is also viewed as an important channel for reducing information constraints, thus providing interested stakeholders with the necessary information when making decisions. This study evaluates one such regime. In particular, we focus on an area of research which has received limited attention to date, namely, the labor market consequences of disclosure for top managers and employees in these organizations.

One potentially important way to assess performance of public sector organizations is via a subjective assessment system. In our setting independent evaluators (inspectors) visit schools at short notice and disclose their findings via publicly available ratings. Although there are an increasing number of papers analyzing the impact of quality assessments and public ratings on the performance of public sector institutions such as schools, evidence on the consequences for public sector employees of such assessment regimes is much more limited.[1]

Furthermore, evaluating the impact of subjective evaluations is important and novel because public sector entities are typically assessed on hard performance metrics (such as the proportion of students attaining critical test score thresholds). Economic theory and empirical evidence suggest that a focus on narrow performance measures leads to strategic employee behavior in a multitasking setting; subjective evaluations have the potential to limit such behaviour (Holmstrom and Milgrom, 1991; Dixit, 2002; Dranove et al., 2003; De Janvry et al., 2023). However, subjective evaluations may lead to a new set of concerns, including the potential for influence activities and favoritism (Milgrom

[1] See for example, the literature investigating the impact of policies such as 'No Child Left Behind' in the US on school resources and organization (Figlio and Winicki, 2005; Dee et al., 2010, 2013; Li, 2015; Feng et al., 2018) and student outcomes (Richards and Sheu, 1992; Ladd, 1999; Greene et al., 2001; Carnoy and Loeb, 2002; Hanushek and Raymond, 2005; Jacob, 2005; Deming et al., 2016; Mansfield and Slichter, 2021).

and Roberts, 1988, Prendergast and Topel, 1996). Yet another concern is the potential for 'ratings compression' (Prendergast, 1999), meaning that evaluators are reluctant to give poor evaluations to low performers. Arguably neither of these issues is a serious concern in our setting given the fact that evaluators typically arrive at the school on a rota system in the period under study, are unrelated to the school or its staff, and 'say it like it is' (for details, see Hussain, 2015). Thus we are able to offer evidence on the effectiveness of a subjective evaluation system which is not subject to these limitations.

In this paper we examine the labor market consequences of subjective school ratings produced by independent inspectors for principals, senior managers and teachers in English state secondary schools. The English school inspection system, undertaken by the Office for Standards in Education, or 'Ofsted', has been in place since the early 1990s. The aim of school inspections is to provide a public assessment of both primary and secondary schools. They involve both an evaluation of the school's recent performance (e.g. share of students reaching acceptable thresholds in key examinations) and on-site inspection with monitoring of classes and interviews with the management and parents. We restrict our attention to state-run secondary schools and to Ofsted inspections taking place between 2006 and 2008. This period is characterized by a uniform inspection procedure, and a reform which dramatically reduced the notice period for schools on the exact date of inspection from many months to just a few days. This resulted in stiffer evaluation criteria.

Using data on the universe of teachers in the English public (state) school system, we compare principals, school managers (deputy and assistant principals, whom we characterize as 'senior leaders') and teachers experiencing a grade change in the inspection rating at their schools with similar individuals in other schools experiencing no change in the rating in a difference-in-differences framework. As discussed in greater detail below, over our period of analysis, timing of inspections is unrelated to performance or other characteristics of the school. We first focus the analysis on principals and then extend it to other school employees. Schools experiencing different outcome of their inspection and school personnel in these schools exhibit similar characteristics in

the period preceding the inspection and the common trends assumption holds for our outcomes of interest prior to the inspections.

Given this setup, we are able to tackle important questions relating to how performance ratings might differentially impact principals versus other managers as well as regular teachers. In particular, as well as the overall school rating, inspectors also provide more granular evaluations or ratings, including one for the overall quality of leadership and management, and another for overall teaching quality in the school. This unique feature of our setting allows us to speak to the precise mechanisms by which ratings impact the labor market.

Furthermore, ex ante, it can be argued that the overall rating provides information relating to the underlying quality of the school, which may not be revealed fully by objective measures such as the school's test score performance. Alternatively, ratings may be valued in and of themselves, and hence the mechanism for the impact of ratings is along a pay for performance channel. For example, it has been shown that an improved inspection outcome leads to greater parental demand (Greaves et al., 2023), which in turn boosts the school's budget via student enrolment. Thus, improved ratings are a valuable outcome for the school, independent of any information about the quality of the principal they may generate for the school's governing board or competitor schools. As explained in detail below, we discuss how these separate channels may be teased apart in our setting.

We find that inspection ratings have important effects on wages, career progression, exit and retirement decisions for school managers, and most notably for principals. A unit improvement in their school rating leads to a 3% increase in principals' salary. It also has dramatic effects on their rate of exit from the (public school) teacher labor market: a unit *increase* in the school's rating leads to a 6 percentage point *decline* in the probability of exit 3 years after the inspection. Around half of this impact on the exit rate can be accounted for by principals' choices regarding early retirement. Furthermore, our results demonstrate that all of these effects are highly asymmetric,

with much stronger impact for schools that experience a decline in their rating compared to schools that experience an improvement. This suggests that governing boards and the market react more by "punishing" deterioration in school quality rather than through rewarding improvements.

For senior leaders and regular teachers, the results go in the same direction, but the overall magnitudes are much more muted. For example, for senior leaders a rise in the inspection rating leads to 1.2% increase in their salary after three years, whilst for teachers the impact is just 0.2% three years out from inspection. We find that senior leaders in schools experiencing an improvement in their inspection rating experience a higher probability of becoming principals.

We contribute to the existing literature in a number of ways. First, we contribute to the literature on the consequences of subjective evaluations in the labor market. As noted by previous researchers (Lazear and Oyer, 2012; De Janvry et al., 2023) empirical evidence on the impact of subjective evaluations is very thin and largely confined to private sector settings (Chevalier and Ellison, 1999; Hayes and Schaefer, 2000).[2] De Janvry et al. (2023) provide experimental evidence evaluating the role of influence activity arising between supervisors and workers in the Chinese civil service. As far as we are aware, ours is the first study to investigate the impact of subjective performance evaluations on school principals and educators. Uniquely in our setting, the evaluations of the quality of leadership and management generated by the inspection process also allows us to unpack their impact on principals *over and above* the overall school inspection rating as well as other performance metrics such as student test scores. We do so in a comprehensive manner, investigating not just turnover, but also the consequences for wages, career progression as well as exit from the teaching profession. We are not aware of any previous study exploiting disclosure of such measures of leadership and management quality and their labor market consequences for top leaders in the education sector.

Our study is also related to a small literature investigating the impact of school ac-

[2]For large private corporations Frederiksen et al. (2017) provide evidence suggesting that subjective performance ratings are correlated with employee career outcomes.

countability policies on the teacher labor market. A handful of studies in this area have investigated the impact of school accountability regimes on teachers (Loeb and Cunha, 2007; Feng et al., 2018; Dizon-Ross, 2020; Gjefsen and Gunnes, 2020).[3] We are aware of two other studies that have investigated the causal impact of accountability measures on school principals, Cullen et al. (2021) and Besley et al. (2021). Cullen et al. (2021) demonstrate that principals just below an unacceptable-acceptable accountability threshold experience negative outcomes in terms of retention and salary, relative to principals just above the threshold. Besley et al. (2021) relate schools' test score performance to principal salaries. In particular, they demonstrate that a reform which increased salary flexibility led to a tightening of the relationship between performance and salaries.

We also contribute to the literature related to CEOs and performance pay (Murphy, 1999) as well as the literature related to employer and market learning about worker productivity (Gibbons and Katz, 1991; Schönberg, 2007). In our setting, although the school's own governing board (responsible for pay setting, see details below) may well be aware of the quality of the principal, the wider market may only observe proxies, such as publicly available school-level information on test scores. These may provide very coarse or imprecise indicators of actual principal quality. The public inspection rating, on the other hand, may provide a significantly more precise signal to the wider market. In our setting, which is characterized by a competitive market for principals, we are able to test for wage and career responses to such disclosure.

Finally, our work also contributes to the literature on the effects of school inspections (e.g. Scanlon 1999; Luginbuhl et al. 2009; Hussain 2015; Greaves et al. 2023), although we are not aware of any studies within the economics literature which have investigated

[3] Loeb and Cunha (2007) provide some interesting evidence on how early accountability policies in the US influenced teacher and principal turnover. Feng et al. (2018) demonstrate that there is little impact of the release of school ratings (based on school test score performance) on teacher mobility, but that mobility of teachers is more likely to raise student test scores at worst performing schools. Dizon-Ross (2020) finds that lower accountability grades are associated with lower teacher turnover and increase in the quality of teachers joining the school. Gjefsen and Gunnes (2020) investigate the impact of disclosing school value added rankings on teacher turnover, finding that the reform led to substantially higher teacher mobility and that teachers leaving were replaced by teachers with higher value added on test scores.

the impact on teachers' careers.

The rest of the paper is structured as follows: Section 2 lays out a simple Bayesian learning model which aids in understanding exactly why the market responds to inspection ratings; Section 3 describes the institutional setting and the data; Section 4 describes our empirical strategy; Section 5 reports our main results; Section **??** investigates heterogeneity across principals and mechanisms for the impact of change in grades, and finally Section 8 concludes.

2 Theoretical Framework

In this section we adopt a standard Bayesian learning model in order to better understand how the market might respond to disclosure of school quality ratings. We make the assumption that the rating provides a noisy signal of the quality or productivity of the school principal. The true quality of the principal, η, is not known, but it is drawn from a normal distribution with a known mean, $\bar{\eta}$, and variance, σ_η^2 :

$$\eta \sim N(\bar{\eta}, \sigma_\eta^2)$$

The precision of the prior is given by $h_p = \frac{1}{\sigma_\eta^2}$. Inspection ratings, $\tilde{\eta}$ provide a noisy signal of true quality:

$$\tilde{\eta} = \eta + \varepsilon$$

where ε is drawn from a normal distribution with known variance:

$$\varepsilon \sim N(0, \sigma_{\varepsilon}^2)$$

The precision of the signal is defined as $h_s = \frac{1}{\sigma_{\varepsilon}^2}$. Given this setup, it can be shown (see, for example, Schervish and DeGroot 2012) that the posterior is given by:

$$E(\eta|\tilde{\eta}) = \frac{h_p\bar{\eta} + h_s\tilde{\eta}}{h_p + h_s}$$

This can be restated as follows:

$$E(\eta|\tilde{\eta}) = \gamma\bar{\eta} + (1-\gamma)\tilde{\eta}$$

where $\gamma = \frac{\sigma^2\varepsilon}{\sigma_{\varepsilon}^2+\sigma_{\eta}^2}$ and $1-\gamma = \frac{\sigma^2\eta}{\sigma_{\varepsilon}^2+\sigma_{\eta}^2}$.

This simple framework helps shed light on why the market might respond to signals of school quality provided by external evaluators. To the extent that the school's own governing board (responsible for pay setting for the principal, within given limits – see details of the institutional setting below) is already aware of the principal's true quality, η, arguably there is little news in the rating (signal, $\tilde{\eta}$). However, there are a number of reasons why the principal's pay may, nevertheless, be affected by inspection ratings. First, the true quality of the principal may only be known (or at least known with a high degree of precision) by the school's own governing board.[4] The public inspection rating therefore provides a signal to the wider market about the quality of the principal. In this case, in an environment where there is a competitive market for principals, we may expect a wage response.

[4]The market will observe other proxies for the quality of the principal, such as publicly available school-level information on test scores. However, these may provide very coarse or imprecise indicators of the quality of the principal.

This framework also helps motivate at least some of the empirical analysis which follows. The response to the inspection rating ($\tilde{\eta}$) will depend on the precision with which the market was already aware of the quality of the principal, as well as the precision of the inspectors' findings. For example, as the technology of inspection evolves and the precision of ratings changes, then the market will take account of these changes when updating priors. In particular, the 2006-2008 inspection reforms entailed very large reductions in notice period provided to schools relating to the precise date of the inspection visit (down from many months to two to three days). If this reduction in notice period leads to a more accurate assessment of underlying school quality (perhaps because there is less scope for 'window dressing' and other gaming behavior) then we can expect that the precision of the signal increases. In this case, the weight attached to the rating can be expected to be higher in the reform period (2006-2008) than in the pre-period. Furthermore, when the variance or noise of the prior assessment is large then, for a given precision of the signal, the market should place greater weight (react more) to the signal. This may be the case for new or recently appointed school principals. A simple contrast in the response to ratings for established versus newly appointed school principals would provide a simple test of this hypothesis.

Finally, although this information-based rationale provides a powerful theoretical motivation for the potential impact of inspection ratings, an alternative reason why the ratings may affect the principal's wages and career outcomes is that ratings may in fact be an *outcome* that the governing board cares about. The ability to deliver good inspection outcomes will be valued by the governing board (for example, because higher ratings may boost parental demand) and hence there may also be an element of performance pay related to inspections, over and above any information imparted by the ratings. In the empirical analysis below, we explain how the rich data and setting that we examine might allow us to shed light on these two distinct channels related to information versus performance pay.

3 Institutional Setting and Data

3.1 The English School System

Compulsory education in state schools[5] in England is organized into five "Key Stages". The Primary phase, from ages 4-11, spans the Foundation Stage to Key Stage 2 (Years 1-6, where Years are the English terminology for Grades). At the end of Key Stage 2, when pupils are aged 10/11, children leave the Primary phase, and go on to Secondary school from ages 11-16. This is divided in Key Stage 3 (Years 7-9) and Key Stage 4 (Year 10-11). At the end of each Key Stage, prior to age-16, pupils are assessed on the basis of standard national tests (though the Key Stage 3 test was discontinued in 2008).

In our study, we consider secondary school personnel in state-maintained secondary schools in England. We focus mainly on principals but also extend the analysis to senior leaders (deputy and assistant principals) and teachers. Secondary education is a crucial part of students' career and previous research has shown that a poor performance in the General Certificate of Secondary Education (GCSE) examination, which is the main examination for students at the end of the Key Stage 4, has long lasting consequences on students' career (Machin et al. 2020), both in terms of further education and in early job outcomes. For this reason, the quality of secondary schools is continuously under scrutiny by the government and by parents.

School governing bodies are responsible for holding teachers and principals accountable for their performance. They are comprised of parents, staff members and, in some cases, local authority representatives. A sub-group of the governing body, along with an external advisor, sets new objectives, monitors progress, and makes decisions about remuneration. The school-level governing board is also responsible for hiring the principal, while the principal and other senior leaders are responsible for hiring the rest of the staff in the school.

Nationally set pay scale determined both teacher and principal wages, although this policy

[5]State schools in England account for around 93 percent of the population of students.

changed in 2014 when teacher pay in England has undergone a drastic reform (Burgess et al., 2022). Teachers' pay-scale was set by the government, based on recommendations by the School Teachers' Review Body. This was an independent body that offered recommendations every year to the Secretary of State for Education and the Prime Minister on teachers' pay and conditions for England and Wales, after consultations with teacher unions and associations, organizations representing school governing bodies and local authorities. The recommendations on changes in pay scales reflected the state of school teacher and head teacher demand and supply but also the wider state of the labor market. Each year, the government used these recommendations to update the School Teachers' Pay and Conditions Document (STPCD). This was the document that school governing bodies followed when setting salaries under this regime. Teacher salary increases took the form of movements along spine points determined within schools, according to the government set salary scale ranges. However, there was some room for discretion, especially for managerial positions (principals and senior leaders). Additional points for career advancement and salaries for managers were decided at the school level.

3.2 Ofsted inspections

Our analysis focuses on the assessment of schools by Ofsted, an independent government authority undertaking inspections of both primary and secondary schools, since 1992. Inspections take into account many aspects of school activity, ranging from the overall performance of students, the effectiveness of classes, the perceived safety of the schools, as well as the quality of management. Inspections usually take place about every three to five years. They can last for a number of days and inspectors typically observe classroom practice, interview students, parents, teachers and the school management (principals, deputy and assistant heads) and review the performance of the school on standardized tests.[6] At the end of the evaluation process, the school is assigned a grade from 1 ("Fail") up to 4 ("Outstanding").[7] Inspection grades are made public in reports available on the

[6] See Hussain (2015) for additional details of the inspection process.

[7] Throughout the analysis we reverse the ordering of official grades. Under the Ofsted rating system the highest grade (4) is given to failing schools while the lowest (1) is assigned to the best schools.

Ofsted website, containing a short description of the outcome of the inspection as well as numerical scores for the main dimensions verified during the inspection. Ofsted scores are heavily based on qualitative information which might not be known to parents and other stakeholders in advance. Hussain (2015) shows that Ofsted scores are related to teaching practices and parents satisfaction beyond what can be predicted by hard test scores, while Hussain (2022) shows that there is a statistically significant relationship between changes in inspection ratings and changes in test performance as well as valued added. Nevertheless, the R-squared in such regressions remains small, suggesting that inspectors likely capture aspects of school quality which are not purely related to test performance. Schools receiving a fail grade are subject to several additional interventions, with follow up inspections over a shorter time horizon (one or two years), managerial interventions and closer monitoring, while no additional measures are implemented for other schools.

In our analysis, we use inspections taking place between 2006 and 2008.[8] This has several advantages. Starting from 2006, inspections became more severe, with stricter standards and a shorter notice period before inspections (which was reduced from many months to just days). This led to a decline in average grade of inspections which was not foreseen by the schools. In the analysis, we use all grade changes, but we also separately estimate the effects of improvements and declines and the impact of specific grade transitions. Due to changes in assessment standard in our period of interest, grade declines are likely to be more unexpected based on previous Ofsted standards.

It is worth mentioning that parents can also access publicly available information on school performance through school Performance Tables. Performance Tables provide summary measures about the performance of students in standardized exams, such as the share of students obtaining 5 or more A* to C in the GCSE, which are the exams at the end of secondary schools, for the overall school and specific groups of students. This is commonly used as a measure of school performance. Performance Tables also provide some information about school characteristics such as number of students, share of

[8] We also extend the analysis to other years and results are in line with our main findings.

students on free school meals, etc. Since 1992, school performance tables are published nationally every year by the Department for Education and are a primary source of information for parents. Since the performance metrics used in Performance Tables fully focus on hard measures of performance, they are unlikely to capture other soft information which are conveyed by the Ofsted inspections such as teaching practices, effectiveness of leadership, safety, etc. We later show how Performance Tables measures relate to Ofsted inspection scores and asses whether they contribute in explaining the impact of Ofsted inspection on principals' career.

3.3 Data

The analysis relies on three main data sources. First, we rely on data on the population of Ofsted ratings for state schools, starting from 1994 up to 2008. Ofsted reports data on the final grade each school received at the end of the inspection, and since 2001 it also reports sub-grades on specific aspects of its inspection such as safeguarding, quality of teaching and quality of leadership. We use inspection grades for schools in the 2006-2008 period, but employ inspection data for earlier periods (pre 2006) to compute the change in rating between inspections. This allows us to see whether schools inspected in the 2006-2008 period experienced a change (improvement or deterioration) in their Ofsted rating. This is crucial as it represents an information shock both to parents and to the school governing board concerning the performance of the school, which we exploit for our identification on the causal effect of inspection on school managers' and teachers' pay and career progression.

To get the information on teachers' pay, career progression, and other characteristics, we employ the Database of Teacher Records (DTR) data, which provides detailed information on school personnel: it contains annual records for school employees collected for pension purposes. The dataset reports information on the broad occupation of school employees and salary. We classify school employees in three main categories: teachers, senior leaders (deputy and assistant heads) and principals. We focus on principals, who are the ones most likely to be held responsible for the performance and management of

the school, but we also report results for senior leaders and teachers.

Finally, We also rely on the Edubase dataset to create a consistent panel of schools. Edubase is an administrative census of the population of schools in England. It provides information on the dates of school openings, closures and school conversions, and some other school characteristics such as stage of education, school location, etc. We consider as new schools those resulting from splits (one school dividing into two or more schools) and mergers (a school resulting from the merger of two or more schools), while we construct a unique identifier for schools experiencing a change in name or other conversions.

We create a unique panel of state secondary school data, matched with teachers and Ofsted inspection data. The panel consists of data from three years before the inspection to three years after the inspection for each school, for both schools and school personnel. We construct a panel for each inspection cohort and then stack them in a single dataset. For the reasons explained above, we restrict our attention to schools being inspected between 2006 and 2008. In this period, about 2,400 schools were inspected out of total of 3,200 state secondary schools. If a school is inspected more than once during this period, we use the change in grade determined by the first inspection. We exclude schools which: (i) did not have a prior Ofsted grade (around 50 schools); (ii) received a "Fail" grade in the 5 years before the inspection between 2006 and 2008; (iii) experienced a grade change in the three years before the inspection;[9] (iv) changed principal between the year prior to the inspection and the year of the inspection, which ensures that the principal is the one affected or 'treated' by the inspection;[10] (v) do not have the entire seven-year period (from three years before the inspection to three years afterwards).

[9]We exclude these schools since these changes might determine a differential trend due to the schools being subject to a special monitoring regime.

[10]This exclusion potentially takes out from the sample extreme cases in which the principal is removed immediately after an inspection. Since we do not have data on the exact date at which a principal leaves or joins the school, it would not be possible for us to exploit more granular information on the timing of the inspection and change of the principal. It is, however, unlikely that intra-year change in principal plays a dramatic role in our analysis since removing an existing principal and finding a replacement takes time. It is also worth pointing out that schools in England experience a relatively high turnover of principals. Besley et al. (2021) show that, on average, English schools have about 10% yearly turnover rate in principals.

Finally, we exclude all principals, senior leaders and teachers who were not observed in the payroll records at least three year before the school inspection. The final dataset contains information on 1,679 schools and Ofsted inspection ratings, 1,677 principals, about 7,250 senior leaders and 74,700 teachers.

We report initial school inspection ratings and the change in ratings for schools in our sample in Table A1. As the table shows, the large majority of schools start with a "Good" rating, with a similar number of schools starting with a lower or a higher rating (note that, as highlighted above, we exclude schools with a baseline "Fail" Ofsted grade). Whilst the Ofsted rating does not change for a large share of schools (about 44%) in the face of stricter inspections in 2006-2008 period, almost the same share(44%) experience a deterioration in their inspection rating. Only a small fraction (about 12%) experience an improvement in their inspection grade. The stricter conditions imposed at the time of the assessment lead to a general deterioration in school grade. This is further testified by Figure A1, which plots the share of schools that experience grade changes in inspections between 2006 and 2008 with respect to inspections between 2003 and 2005 and the share of schools experiencing a downgrade over time. Inspection taking place between 2006 and 2008 show a higher share of downgrade than inspection which took place in the three years before, as shown in Panel (a). This higher share of downgrade in this period appears even starker if we consider the proportion of schools suffering from a downgrade over time. Panel (b) provides a description of the proportion of schools experiencing a deteoriation of their inspection rating since 1995 and it clearly shows a jump from 20% up to 40%, starting from 2006, and remains stable for the following two years.

4 Empirical Framework

We employ difference-in-differences (DID) models to isolate the causal impact of school inspection, where we compare schools experiencing an improvement or decline in their inspection ratings (our treatment group) with schools experiencing no change in their Ofsted ratings (our control group). As discussed earlier, although observable factors such

as school past performance based on national standardized tests can help predict changes in ratings, a very large proportion of the variation in ratings cannot be explained by readily observable characteristics of schools. Arguably then, inspection ratings provide new information or 'news' about school quality, over and above the news provided by school performance tables, which is orthogonal to other publicly available measures of performance. Indeed, as we demonstrate below, schools and principals' outcomes are on parallel trends in the periods before the change in inspection rating, which supports the view that Osfted inspections are providing new information which were not available before.

To identify the causal impact of a change in Ofsted ratings, we estimate the following model:

$$y_{iprst} = \alpha + \beta_1 Post_{rt} + \beta_2 ChangeRating_s + \beta_3 Post_{rt} X ChangeRating_s + \\ + X_{it}\gamma + \delta_{ir} + \lambda_{prt} + \varepsilon_{iprst} \quad (1)$$

where y_{iprst} is the outcome of interest (salary or an indicator for the exit from the school sector) for individual i with tenure p in school s in year t, experiencing a school inspection event in year r ($r \in \{2006, 2007, 2008\}$), - we refer to these years as cohorts throughout the rest of the paper. Tenure is fixed at the time of the inspection. Practically, we compare individuals who experienced a change in Ofsted school rating (treated) and individuals who did not experience a change in Ofsted school rating (comparison group) during this period, conditional on the fact that they had the same tenure at the time of the inspection. $Post_{rt}$ is set to 1 in the years following the inspection, and is zero otherwise. $ChangeRating_s$ is the change in the school inspection rating between the current inspection and the previous inspection, and it can take integers values between -3 and +2.[11] The equation also includes a polynomial in age (X_{it}) and fixed effects at

[11] As discussed above, we exclude +3 since we drop schools previously receiving a "fail" grade, corresponding to 1 on our scale.

the individual-cohort level (δ_{ir}), and at the cohort-tenure-year level (λ_{prt}). Hence, our model estimates the impact of the change in inspection rating by comparing, over time, principals with the same tenure at the time of the inspection, but who experienced different change in Ofsted rating, within the same cohort of inspections (2006-2008) after netting out individual time invariant differences. The treatment status of a principal is determined by the change in rating for the school in which they were at the time of inspections taking place between 2006 and 2008. Standard errors are clustered at the level of school (treatment level) at the time of inspection. The parameter of interest is β_3, the impact of a unit improvement in the school's inspection rating.

To assure that there are no pre-treatment trends, we also present our findings in an event study framework, estimating models of the following form:

$$y_{iprst} = \sum_{\tau=-3;\tau\neq-1}^{\tau=3} \alpha_\tau \mathbf{1}(t-r=\tau) + \sum_{\tau=-3;\tau\neq-1}^{\tau=3} \beta_\tau \mathbf{1}(t-r=\tau) ChangeRating_s + \\ + \xi ChangeRating_s + X_{it}\gamma + \delta_{ir} + \lambda_{prt} + \varepsilon_{iprst} \quad (2)$$

where the indicator function $\mathbf{1}(t-r=\tau)$ is switched to one when the difference between the year (t) and the year of of the inspection cohort (r) equals τ. These dummies represent the distance in years from the year of the inspection by cohort and are defined from -3 years before the inspection up to +3 years after the time of the inspection event. Here, the parameters of interest are the coefficients β_τ: they represent the impact of a unit improvement in the school's inspection rating at event time τ relative to the omitted category, $\tau = -1$, i.e. the year before inspection.

It should be noted that our empirical strategy does not exploit the timing of the inspection for the estimation, but, rather, it compares schools experiencing change in Ofsted rating with schools not experiencing a rating change within the same cohort of inspection (i.e. all schools we compare schools that were inspected in 2006 and experienced a change in score with other schools which were inspected in the same year

but did not experience a Ofsted rating change). Hence, our estimates are not affected by the issues related to staggered DID event study estimators, which have been extensively discussed in recent studies (e.g. Borusyak et al. 2021, Callaway and Sant'Anna 2021)

4.1 Threats to identification

Parallel trends

The key identification assumption underlying our empirical strategy is that the outcome follows the same trend for treated and untreated schools, in absence of the treatment.

In order to test the credibility of the common trends' assumption, we assess whether there is any evidence to suggest that it is violated in the pre-treatment periods. The event study framework provides a natural way to test for this. As we demonstrate in Figure A2 in the Appendix, we find no evidence of such divergence. The figure, which reports estimates for equation 2 for principals' main career outcomes with the sample restricted to the period before the inspection, shows no differential trend across principals which will experience differential changes in Ofsted ratings. This is further confirmed by p-values for a test of the joint equality of coefficients to zero, reported at the bottom of each panel. The test never rejects the null hypothesis of the joint equality of the coefficients for the pre-period to zero. Schools experiencing different rating changes between 2006 and 2008 are on parallel trends for our outcome of interest in the periods before the inspection and this supports our parallel trend assumption.

Timing of Inspections

Another potential concern is that, although it can be demonstrated that in the pre-treatment period, treatment and control groups follow common trends, they may in fact diverge at the point of treatment, even in the absence of treatment (disclosure of ratings). This would be the case if, for example, timing of inspections is endogenous and an inspection takes place in correspondence of positive or negative shocks to the school performance. Under such a scenario, correlated time-varying unobservables drive both the treatment status as well as the outcome. However, such concerns are mitigated

in our setting because school inspections are carried out on an inspection cycle rota, and timing of inspections is unrelated to observable time-varying school characteristics. Table A2 in the Appendix shows a cross-tabulation for the distance in year between the last inspection and inspections taking place between 2006 and 2008. It clearly emerges that the majority of inspections take place at fixed frequencies (89% of schools inspected in 2006 had been inspected 5 or 6 years earlier). This changes over time as the system shifts towards a three to four year cycle but, by 2008, 79% of schools inspected had received their previous inspection 3 or 4 years before.

In addition, observable characteristics play a little role in the timing of the inspection. Table A3 estimates a linear probability model for the probability of receiving an Ofsted inspection and it considers several principal and school characteristics (both in levels and in change over the two years prior to the inspection) as well as variables for the last inspection (distance in year and past grade).[12] Distance from the previous inspection appears to be the most relevant factor in determining when a school is going to be inspected again. Coefficients for dummies for the distance in years from the last inspection are sizable and contribute around a third of the variation in the dependent variable ($R^2 = 0.312$). Past grade also plays a role, but coefficients are substantially smaller and it contributes only for an additional 0.047 in terms of R^2.[13] principal and school characteristics play only a minor role: they are rarely significant and, even when they are, the magnitude of these coefficients is negligible with respect to previous variables. Overall, the available evidence suggest that Ofsted inspections follow a predetermined temporal pattern, with minimal variations, and they do not respond to changes in the school performance.

Comparability of schools experiencing different grade changes

It could also be argued that principals/schools experiencing improvements or declines in Ofsted grades may differ substantially between each other, but also with respect to principals/schools experiencing no change in inspection grade. This would make the latter

[12]The model excludes schools which received a "fail" grade in the past, due to the special interventions they are subject to, and schools with no prior inspection.

[13]Our results also hold conditional on the prior grade. We explore this explicitly in Section 5.4.

group a poor counterfactual for schools experiencing a change in rating. However, these different groups do show remarkable similarities in terms of observable characteristics, thus alleviating such concerns. Table 1 reports summary statistics together with t-statistics for the difference between schools experiencing a change in rating and schools experiencing no rating change. Column (1) reports average characteristics for schools with no change in Ofsted rating, while Column (2) and Column (4) report average characteristics for school experiencing a grade decline and a grade increase, respectively. Finally, Column (3) and Column (5) report t-stats for the t-test for the equality of means between schools experiencing no change in inspection grade and those experiencing a grade decline (Column 3) and between schools experiencing no grade change and a grade improvement (Column 5). Differences across principals/schools experiencing an Ofsted grade improvement, an Ofsted grade decline and no change in an Ofsted grade are generally negligible even when statistically significant (see for example tenure for teachers, where, for both the difference between no change and decline and no change and improvement, the difference amounts to less than half a year in tenure). There are some differences in terms of school performance with schools experiencing a grade decline generally performing a bit worse than schools experiencing no change (the reverse is true for schools experiencing a grade improvement). This is expected since Ofsted grades should reflect, at least partially, school performance. However, this is not detrimental for our identification strategy since this does not generate differential pre-treatment trends in our outcomes.

5 Results for School Principals

5.1 Salaries

We first provide some simple graphical evidence on the impact of inspection ratings on salaries for principals. Figure 1 simply plots log salary in and around the event time, normalized at the year before the inspection $\tau = -1$, for schools experiencing a rating improvement (dots), a decline (square) or no change (triangle) in their inspection rating.

The figure provides suggestive evidence that there is minimal divergence in trends for the three categories of schools before treatment, whilst in the post-period we see a clear divergence: principals, whose schools experience a decline in Ofsted grade have lower wages, while principals experiencing a grade increase get, temporarily, higher wages. Reassuringly, principals of schools experiencing no rating change remain on the same trend they had before the inspection.

We then quantitatively assess the initial graphical findings in our DID regression framework and report estimates of model 1 in Table 2 for (log) salaries. Column 1, which only includes the difference-in-differences term and a polynomial in the age of the principal, shows that the impact of a unit improvement in the school's rating leads to a 1.7 percent rise in salary for principals, on average, over the three years following the inspection. Columns from (2) to (4) extend the regression model by progressively including richer sets of fixed effects. More specifically, Column (2) includes individual-by-cohort,[14] year, and cohort (year of inspection) fixed effects. Column (3) adds tenure as a principal at the time of the inspection fixed effects, cohort by year, tenure by cohort, and tenure by year fixed effects. Finally, Column (4) includes a triple interaction between tenure at the time of inspection, cohort, and year fixed effects. Hence, in the this last specification, we compare, over time, principals who were subject to an inspection in the same year and had the same tenure at the time of the inspection. This is our preferred specification for the rest of the paper. It is worth pointing out that the coefficient of interest remains stable and the richest model (Column 4), which includes a large set of fixed effects, provides estimates that are extremely close to our baseline equation in Column (1). Finally, Columns (5) and (6) use salary in levels as an outcome, with Column (6) imputing zeros when the principal is no longer observed in our dataset. The latter provides an overall impact of an Ofsted grade change, under the assumption that principals leaving the profession do not get a wage outside the school system, and it provides suggestive evidence on the impact of inspection ratings beyond wages.[15]

[14]The cohort dimension (i.e. the year of inspection) matters very little since only a few individuals appear as principals in more than one school in the sample.

[15]Due to lack of data on wages outside the school sector, we are unable to discuss what happens to principals leaving the school sector, although we show later that many of them retire and are unlikely

Column (5) shows that an improvement in the Ofsted rating by one grade leads to a salary increase close to £1,500 per year. The effect becomes much larger when we account for exit in our model by imputing zeros for principals no longer observed in our data. Column (6) suggests that turnover plays an important role in this setting and we investigate this important margin in the sections below.

Finally, Figure 2 plots the event study estimates of model 2 for principals. It demonstrates that in the years after experiencing a unit improvement in their school's rating, the difference in principal's salary (compared to principals who experience no change in their school rating) steadily rises, up to 3% per point in grade improvement, three years after inspection. Importantly, years prior to the inspection do not show any differential trend between the treatment and the control groups ($\tau = -2, -3$), in line with the results reported in section 4.1. This supports our identification framework and shows that, despite some small differences in levels in the performance of the school, principals were on similar trends in outcomes prior to the additional signal provided by the Ofsted inspection. The absence of any pre-trends provides evidence in favor of the causal interpretation of our findings.

Heterogeneity

We now assess how the impact inspection ratings varies by principal characteristics. We start by considering the role of principals' characteristics such as gender, age, and tenure. First, we look at the gender dimension. There is established evidence on the disadvantages faced by women in the labor market and it is possible that this extends also to women in managerial positions in the public sector. Women are generally less likely to have managerial positions in this period and only about a third of principals in our period are women, although the share had been increasing over the years (from 20% in 1994 to close to 32% in 2006). In addition, women have been shown to be less likely to bargain with their employers, and this might expose them to the decisions of the school governing board after the disclosure of the inspection grade (Hernandez-Arenaz and Iriberri, 2019). If a principal is less willing to bargain for her position or defend her

to have other employment.

work after a negative review from Ofsted, it might be easier for the governing board to implement harsher penalties against her. Second, we consider age. Individuals at different ages have different opportunities on the labour market, and this could also lead to differences on how their career is affected by the Ofsted inspection. Younger individuals, for example, are unlikely to retire and may remain in their job even if they suffer larger wage penalties from their school if the latter performs badly in Ofsted inspections. Age, consequently, might determine which margin of the career of the principal, extensive or intensive, is mostly responsive after the Ofsted inspection. Finally, we look at tenure at the time of the inspection. This dimension should play an important role for the impact of inspections: a principal who has just been hired is less likely to be associated with the change in the performance of the school, while this is not the case for an individual who spent a long time in the school. We now consider separately all these dimensions on the analysis.

We estimate our main model on different subgroups of individuals and report results in Table 3. Column (1) reports baseline estimates for reference while Columns (2) and (3) compare men and women principals. Women do not seem to be affected more than men. On the contrary, women, generally, seem less exposed to consequences of changes in Ofsted rating, especially in terms of their pay (+2.3% per grade change in Ofsted rating for men and +0.6% per grade change in Ofsted rating for women) while they are still affected in terms of exit probability but less than men (-4.5% per grade change in Ofsted rating for men and -3.3% per grade change in Ofsted rating for women).

We move on to assess differences between older and younger principals, where we define 'older' principals the ones with a higher than median age in the year before the inspection (52 years). We report results in Column (4) and (5) of Table 3, respectively. A point improvement in Ofsted grade does not affect the wage of older principals, but it leads to higher wages (+2.8%) for younger ones. However, older principals see a much larger decline in the probability of exiting the school sector for each grade point increase in Ofsted rating (-6.6%) with respect to younger ones (-2.5%). Since a large share of principals leaving the school sector as a consequence of Ofsted rating changes retire, as

shown in 3, the stronger impact on exit for older principals appears reasonable. Hence, it appears, that the age of the individual and the stage of her career plays an important role in how the the impact of changes in Ofsted scores unfolds: younger principals see both higher salaries and a lower probability of exit while the impact for older principals is fully concentrated on the extensive margin, with negligible adjustments on wages.

Finally we consider the role of tenure. We compare principals with less than three years of tenure in the year of the inspection (which corresponds to the bottom quartile of the distribution), with those in higher quartiles. One may argue that a principal who spent many years at the school could be considered more responsible for her school quality than a principal who joined the school just a few years before inspection. In the latter case, the change in grade is more likely to be perceived as the legacy of her predecessor and it should be less consequential for the principal. Columns (6) and (7) of Table 3 report results for principals whose school tenure was more or less than three years at the time of the inspection. Results are in line with our expectations: principals with higher tenure at the time of the inspection are the ones most affected by changes in Ofsted grades, both in terms of salaries (+2% vs +1.3%) and in terms of probability of exiting the school (-5.2 vs -2.2 percentage points).

Taken together, these results suggest that the characteristics of the principal, most importantly age and tenure, play an important role in how the Ofsted rating change affects their career, with larger effects on salaries for younger heads and larger effects on probability of exit for older principals. In addition, we generally find larger effects for principals who already spent several years in the school in their role. Gender, on the other hand, does not seem to be very relevant.

5.2 Robustness Checks

One possible concern is that, while trends close to the inspection are similar, longer term trends might show some divergence and performance might follow cycles connected to inspection timing. To assess this possibility, we further extend our pre-period to the full

inspection cycle (i.e. up to 5 years between two inspections) prior to the changes taking places between 2006 and 2008. Results, reported in Figure A3 of the Appendix, show that even years farther away from the year of the inspection show patterns consistent with the trend in the three years preceding the inspection. Also in this case, we see that the difference between the two groups is flat and in no case there is a statistically significant deviation with respect to the baseline period.

Second, in our main analysis, we limit ourselves to a relatively short period of time over which inspections took place (years between 2006 and 2008). This choice is driven by the presence of a uniform and stricter inspection framework and the possibility to use additional indicators (sub-score of the Ofsted inspection on different aspect of the performance of the school), which help us investigate how Ofsted inspections impact the principals career (see Section 6.3). It is not *a priori* clear if our results would apply to Ofsted inspections in other periods of time. To address this potential concern, we extend our empirical analysis to other years as well. Thus, we construct a school panel for all schools subject to inspection between 1998 and 2008 to assess the validity of our findings over a longer period of time. Similarly to the main analysis, we construct a panel for the three years before and the three years after each inspection and then stack the panels for each school. We then estimate our event study difference-in-differences across the whole period and report the related coefficients in Figure A4 in the Appendix. Again, for this much larger sample, we find a strongly statistically significant effect for a one grade improvement in Ofsted rating, which shows that our results are valid and, actually, even stronger when other inspection cycles are considered.

Finally, it is possible that, despite the many elements unobserved to parents and other schools, agents are still able to predict, to some extent, the change in the rating of the school on the basis of recent observable characteristics. In order to assess whether this could be driving our results, we construct a measure of unexpected change in the rating of the school by using an ordered probit model to predict the grade a school will receive during an inspection based on past characteristics and dynamics in terms

of performance, similarly to Greaves et al. (2023).[16] Once we predict the rating, we compute the unexpected change by subtracting the predicted score from the realized Ofsted rating. We then replace this measure in our equation 1. Results, reported in Table A4 in the Appendix, show that this exercise yields results which are extremely close to our main estimates, and the impact of a inspection grade change is actually the same when we consider our preferred specification in Column (4). This further corroborates our results and makes it unlikely that they are driven by expected change in Ofsted ratings.

Related to the previous point point, it might also be possible that, over time, parents and the school governing board improve their ability to predict the changes in Ofsted ratings and the effect actually fades away over time, so that, in the new equilibrium, rating changes are fully internalized *ex ante* and no longer impact principals' remuneration. If this is the case, then we would expect the impact of Ofsted to decline over time as agents get used to the new criteria. We asses whether this is the case by estimating the impact of a grade increase by cohort of inspections and plot them in Figure A5 in the Appendix. The pattern of the effect is inconsistent with this kind of learning, since the impact of a grade improvement is the largest for the last cohort of inspections. It is insignificant only in 2007, but a test for the equality of the effects (p-value reported below the Figure) cannot reject that the three coefficients are the same. This indicates that parents and governing boards are not able to predict changes in Ofsted rating, even after some time that the new inspection system has been in place, and change in ratings still influence the career of principals.

5.3 Turnover, Career and Other Margins

We now assess the impact of changes in inspection ratings on other dimensions of principals' career, namely whether the principal still holds onto the job after the

[16]The ordered probit considers the following explanatory variables: number of pupils, share of free school meal eligible kids, share of students on special education needs (with certificate), pupil teacher ratio, share of students with 5+ A*-C in the GCSE, share of students with no passes and changes in these variables between two years before and the year before the inspection and the the previous Ofsted rating.

inspection, leaves the school or exits from the school sector. The school governing board might react to a rating improvement by putting efforts to retain effective managers while, conversely, they might remove them in case of unsatisfactory Ofsted performance. In the latter case, managers might move to another school in the public system or exit the school sector, by either retiring or changing sector.

To investigate these margins, we refer to our model described in Equation 2 and estimate linear probability models for our outcomes of interest: probability of being a principal, changing school with respect to the school they were in $\tau = -1$, or exiting the school sector and retiring. We report the results in Figure 3. Our findings show that Ofsted inspection ratings have a strong impact on principals' career. A one unit increase in ratings leads to a 6 percentage points increase in the probability that the principal remains in the role three years after the inspection (Panel a). This is mirrored by a decline in the probability of leaving the school in which s/he was in the year of the inspection (Panel b), and exit from the school sector (Panel c), which both decline by about 6 percentage points after three years since the inspection. Smaller effects, but still sizable, can be detected on the probability of retirement (Panel d), which declines by 3 percentage points over the same time horizon. Thus, about 50% of the principals' exits is driven by a higher retirement probability. These large effects confirm the insight from Table 2 which suggested an important role for extensive margin responses: the probability of leaving the school sector is heavily affected by changes in Ofsted rating.[17]

5.4 Asymmetric Impact by Type of Rating Change

In order to test whether positive or negative inspection outcomes matter more for the principal's career, we investigate separately, the impact of rating improvements and declines on wages. To do so, we define two different samples: the first one consists of schools experiencing grade improvements (treatment group) and schools experiencing no change in rating (control group); and the second sample includes schools experiencing a

[17] One caveat to our analysis here is that we do not observe moves to private sector schools. However, given the widespread dissemination of inspection ratings in this setting, it seems unlikely that principals from under-performing schools would be readily hired by private sector schools.

decline in rating (treatment group) and school experiencing no change (control group). We further decompose the effect of a rating decline into the effect of a decline in grade which leads to a "Fail" grade for the school and a rating decline which does not. Results, reported in Figure 4, show that while a grade improvement leads, at best, to a temporary increase in principal's pay (Panel a), a grade decline has long lasting consequences for her (Panel b) with effects as large as 4% of the salary after three years. This suggests that deterioration in school quality are much more salient for the remuneration of school top managers. These effects are larger for a rating decline which ends up with a "Fail" grade for the school (Panel c) with respect to other rating declines (Panel d), especially in the short run. Three years after inspection, the effects are comparable for the two types of grade declines and close to 4% of the salary.

We repeat this analysis, but we examine the impact on principal's exit from the school sector instead of the remuneration. Figure A6 in the Appendix decomposes this effect into the impact on exit when a school experiences a rating improvement and the impact on exit when it experiences a rating decline. Results show that the impact is much larger for principals in schools that experience rating declines (Panel b) than principals in schools that experience rating improvements (Panel a). The effect is particularly large when schools receive a "Fail" inspection grade. For schools experiencing a "Fail" rating, the probability of the principal exiting the teacher labor market is around 12 percentage points higher three years after the inspection, relative to principals experiencing no change in their rating (Panel c). The impact is more contained for rating declines which do not lead to a "fail" grade (with a cumulative impact of 4 percentage points after three years), as shown in Panel (d). These results confirm that, generally, "punishments" play a stronger role than rewards and the labor market consequences appear to be particularly salient in case a school receives a "Fail" grade.

We further explore the heterogeneous impact of declines and improvements in inspection ratings by conditioning on the prior grade. Indeed, the impact of a grade change might depend on the position from which the school is starting from and a decline in the rating could be perceived more negatively if the school is already struggling. Practically, we

select all schools starting with a certain grade and then compare the career of principals in schools experiencing a grade decline with those in schools experiencing no change in their grade. We report results for "Fail" and "Not Fail" grades in Figure A7 (for salaries) and A8 (for Exit) in the Appendix. We exclude grade improvements from this analysis for simplicity and report related results in Figure A9 in the Appendix.

We start by looking at salaries in Figure A7. If a school experiences decline in rating without reaching the "Fail" grade, then the principal suffers a wage loss close to 5% of her salary three years after inspection. Note that we exclude the baseline grade 2,"Satisfactory", since a grade decline would lead automatically to a "Fail". A "Fail" instead leads to much larger impact for schools with already a low grade; a similar impact for schools with a medium grade; and very small or no impact for schools with very high initial grades. It should be noted that these results are conditioned on the principal remaining in the state school sector. Furthermore, this is the impact of a decline of one grade. The latter element implies that the rating decline to "Fail" for schools with an initial "Good" grade, leads to a 10% salary loss, since they have to go down by two grades, which is comparable to the one registered by schools with lower grade ("Satisfactory").

Next, we consider exit from the school sector for both types of grade changes in Figure A8. For a non-fail rating decline, a grade change is associated with a 10 percentage points increase in the exit probability for both schools with an "Outstanding" and "Good" initial grade. Instead, when "Fail" is considered, the impact is generally larger, with a 17 percentage point higher probability of exit per grade change for schools starting with an "Outstanding" grade after three years and a 15 percentage point higher probability per grade change for schools starting from a "Good" grade. This leads to an overall impact of more than 45 percentage points for schools starting from the top grade (they lose three grades to get to a "Fail", from 4 to 1) and about 30 percentage points for schools initially classified as "Good" (they lose two grades to get a "Fail", from 3 to 1). The large impact on exit for principals in schools originally rated as "Outstanding" suggests that, for such large variations in the school assessment, principals suffer massive

reputation losses and are unlikely to work in English state secondary schools in the future. Hence, it is not possible to observe any impact on salaries. The effect is smaller for schools starting a "Satisfactory" initial grade, with a 10 percentage points higher probability of exit (not statistically significant at 5%) after three years.[18]

Finally, we move to results for an improvement in Ofsted rating, reported in Figure A9. Inspection rating improvements do not determine any change in the salary and lead to a decline in the probability of exiting from the school sector (about 6 percentage points three years after the Ofsted inspection). The effects on exit are, however, imprecisely estimated and the impact is significantly different from zero at 5% after three years only for schools starting from a "Satisfactory" grade. For shorter time horizons after the inspections, the effects are smaller and never statistically different from zero at 5%.

All in all, the above evidence suggests that Ofsted inspections can have important consequences for principals. Experiencing a grade decline has significant impact on salaries and on the probability of the principal retaining their job and this is not compensated by employment in other schools. The stigma appears to be long lasting and leaves a deep impression in the market. Many of the principals respond by retiring.

6 Potential Mechanisms: Information, Pay for Performance and Competitive Pressures

There are a number of hypotheses which may explain why disclosure of inspection ratings impacts the careers of school principals. First, the governing board may not care about the overall school rating itself but rather care about what a change in rating implies for the enrollment at the school and for the characteristics of the student body. Indeed, a lower grade may lead students to leave the school or reduce new enrollments, which may lead to lower funding for the school and higher probability of closure int he future. In this case, the board may be less concerned about the quality or the performance of the principal *per se*, but place greater weight on how inspections ratings affect future

[18]Note that for these schools the grade change required to get a "Fail" grade is only one point.

developments at the school. on the ability of the principal to deliver improved inspection outcomes. That is to say, an inspection rating may in fact be viewed as a performance metric in itself, and hence the governing board may use it as a pay for performance mechanism. Second, although public disclosure of school quality ratings may in fact not reveal any new information to the school's own governing board, the ratings may well be 'news' for other actors, such as rival school boards, who may use these in their external recruitment decisions. In this case, the response to ratings may depend on the level of competition between schools in the principal labor market. Finally, the change in ratings may provide new information to the school governing board itself on the quality or performance of the principal. The board can then use this signal to reward or penalize the principal. In this sense the more tightly the performance of the principal is connected to the rating, the stronger will be the impact of a change in rating.

In this section we shed light on the significance of each of these different channels. We do so by proceeding in the following way. We first demonstrate that ratings have consequences, over and above other publicly available school performance metrics, along margins which matter to the school's stakeholders, including enrollment and the socioeconomic mix of students (which can reflect peer quality, an attribute typically valued by parents).

Then we turn to how competition relates to the impact of changes in inspection rating and consider how the impact of Ofsted rating change varies depending on the number of school at different distances.

Finally, in the last subsection below, we exploit sub-ratings, including one on the quality and effectiveness of leadership and management at the school, in order to provide a direct test of whether the governing board values the overall ratings versus those related to leadership and management.

6.1 Impact of Ratings on Enrollment and Composition of the Student Body

The change in the inspection rating could generate changes in the characteristics of the school, which are relevant for the school governing board. An improvement in the inspection rating could lead to, for example, more students enrolling in the school. Such positive effects of the ratings may be valued by the governing board, and hence they may reward the principal accordingly.

In this section we examine how school outcomes evolve as a consequence of inspection rating changes. We obtain data from the school performance tables on a number of measures such as: the number of students; the share of students from a disadvantaged background (i.e. eligible to free school meals, fsm); the share of students with special education needs (SEN); the share of students receiving 5 or more A* to C grades in the final exams of secondary school (the GCSE); and the number of students with no passes in the GCSE exams. A change in the composition of the student body could lead to higher wages and this channel may explain the wage dynamics observed after the school inspection. Conversely, if the school was to suffer a decline in performance after a positive rating, due to mean reversion, for example, the positive impact of a grade change might be attenuated by a lower pay induced by a worse performance of the school after the inspection. Indeed, information on students' performance in the final exams of secondary school is readily available to parents and constantly scrutinized due to the high stakes nature of the exam. Consequently, variation in these measures of performance could lead to changes in salaries.

We start by looking at how the change in inspection grades affects school outcomes. To do so, we perform our difference-in-differences estimation as described in Equation 1, but we consider the school the unit of the analysis and follow it from three years prior to three years after the Ofsted inspection. We plot results of our event study difference-in-differences estimates in Figure A10 in the Appendix. The results demonstrate that a unit increase in the inspection rating leads to a 3.8 percentage point increase in the

number of students after three years. The composition of the student body also changes with a decline in the share of students from disadvantaged backgrounds, and an increase in the share of students with special education needs.

In terms of school performance we see a decline in the share of students obtaining 5 or more A*-C in the GCSE exams and an increase in the share of students with no passes. This is consistent with previous evidence from Greaves et al. (2023). The trend in the years before the inspection for the share of students achieving 5 or more A*-C in the GCSE exams is not surprising, since the past performance of the school is one of the main indicators used by inspectors for their evaluation of the school. Indeed, we would expect schools experiencing a positive grade change to be on a positive trend in terms of performance in the years prior to the inspection. Interestingly, variations in this continuous measure of school GCSE performance, do not lead to changes in principals' salaries as shown by parallel trends for salaries in the periods before the release of the Ofsted inspection report (Figure 2). It should be noted that the change in share of students achieving 5 or more A*-C is rather small. Indeed, in 2006 schools on average had 58% of students achieving 5 or more A*-C in the GCSE, which implies that the variation over two years corresponds to less than 3% of the average performance of the schools. Such small changes might not be fully appreciated by parents and the school governing board.

Since the inspection rating appears to affect important dimensions of schools, such as the number of students, it is possible that at least part of the impact on salaries is mediated by these effects. To assess how variations in school size, characteristics of the student body, and the performance of students contribute to the changes in principal pay, we include in our regression model school performance indicators and student body characteristics described above (i.e. number of students, share of students eligible to free school meals, share of students with special education needs, share of students obtaining no passes in the GCSE exams, and the share of students attaining 5 or more A*-C grades in GCSE subjects exams). The inclusion of student characteristics and performance, allows us to investigate to what extent these margins are relevant in the

impact of inspection ratings on principals' career.[19] We report results in Column (2) of Table 4. Similarly to the previous Table, Column (1) reports baseline estimates for comparison. The coefficient of this augmented equation is very close to the original coefficient (+ 1.8% in the baseline equation vs +1.6% when school characteristics are included). This suggests that changes in the characteristics of the school explain, at best, a very small share of the effect of the change in Ofsted grades.

6.2 Competition Effects

The environment in which the school operates might play an important role in determining how the impact of ratings is mediated. In particular a more competitive environment could lead the governing board to place a stronger value on the school quality assessment since parents can easily enroll/move students to other schools if the school quality worsens. This would have consequences for school funding (which is in part determined by the number of students enrolled at the school) as well as the prestige of the school.

The presence of greater competition in the local area could lead schools to take more drastic measures if their school quality deteriorated. Indeed, since school funding is, in part, determined by the number of students, loss of students to competitor schools would generate losses for schools which might prompt their governing board to take stronger action in response to changes in the school evaluation. To check if this is the case, we proxy competition by the number of other secondary schools within a certain distance, and ask how this affects the response to changes in their evaluations. We divide schools in two groups based on the number of students within a 5 mile radius and use the median number to define the two groups of schools (6 schools).[20] We also

[19] Note that the coefficient of interest cannot be computed when exit is the dependent variable: school characteristics can be defined only when the principal is in a school and so, in all cases, when exit is equal to one, it is not possible to define the additional variables relating to students and their performance and to estimate the impact of exit after netting out these factors.

[20] According to the National Travel Survey, in 2006 secondary school students traveled for about 3 miles to get to school (see https://www.gov.uk/government/statistical-data-sets/nts06-age-gender-and-modal-breakdown#school-travel), so we might expect students to consider mostly travel within 5 miles to reach their school.

consider the number of schools within a narrower radius (2 miles, one school is the median) and a larger radius (10 miles, seventeen schools is the median).

Columns (3) and (4) in Table 4 investigate the impact of local competition by splitting the sample depending on the number of other schools within a 5 miles radius. Results show that the impact of a grade change in inspection rating is not stronger when there are more neighboring schools. The impact appears to be larger when there are actually fewer schools within 5 miles, as reported in Column (3), although the estimates are rather noisy and the difference in magnitude is not large. This is true for both the impact on salary (Panel a) and for the impact on exit (panel b).

We also estimate the impact of change in inspection ratings by considering different distances and reporting the corresponding impact of grade changes in Figure A11 in the Appendix. As can be seen from this figure, the number of neighboring schools by distance does not seem to play a major role in the relative response to inspection rating changes.

Overall, the analysis in this section suggests that competition, at least as measured by the number of schools in the local schooling market, does not seem to play a major role in mediating the impact of inspection ratings.

6.3 Granular Inspection Ratings: Leadership, Teaching Quality and Student Achievement

So far we have considered only the overall school inspection grade. However, as well as the overall rating, inspectors also provide more detailed ratings for various margins of school quality, including one for the quality and effectiveness of the leadership and management, as well as ones for overall teaching quality and student achievement. In this section we use these subcomponents of the ratings in order to provide a direct test of whether the governing board places greater weight on the overall rating versus that placed on the ratings for management and other subcomponents.

The three subcomponents or sub-grades on quality of management, overall teaching

quality and student achievement are available from 2001 onwards.[21] To investigate the relative role of different aspects of the school evaluation, we replicate our main difference in difference model by replacing the change in overall Ofsted grades, which we used in previous estimates, with changes in the three Ofsted sub-grades listed above, one at a time. We also estimate our difference-in-differences model by including in the regression the changes for all the sub-grades indicators at the same time. Results are reported in Table 5. It should be noted that information on changes in sub-grades is available only for a subset of the inspections between 2006 and 2008: since we have data for subscores only after 2001 and some schools inspected in 2006 had been inspected earlier than 2001, it is not possible to include these schools in the analysis. Hence, as a first step of our analysis, we replicate our main model in Column (1) and then look at separate sub-grades in the following Columns. In Column (2) we replace the change in overall Ofsted grade with the change in grade for pupil achievement, Column (3) considers teaching quality, and Column (4) uses quality of the leadership at the school. Finally, Columns (5) and (6) include all the three measures together. Column (5) also includes the change in the overall Ofsted score, while Column (6) includes only the change in sub-grades. We look first at salaries' in Panel (a) and then to the probability of exit in Panel (b).

Column (1) of Panel (a) in Table 5 shows that the baseline difference-in-differences results based on this sample are extremely close to the ones of the regressions based on the whole sample, so the small loss of schools in this sample has no impact on our main results. Moving on to the impact of the sub-grades on salary in Columns (2) to (4), we find that the items which matter the most appear to be the quality of teaching and the achievement of pupils, which have arguably a stronger connection with the general performance of the school. The change in the quality of leadership appears to have a lower impact with respect to the other measures. Interestingly, if all aspects of the inspection outcome are jointly considered in Column (5), sub-grades for teaching quality

[21] Due to changes in the exact dimensions of the schools examined during the inspection, we focus here on a small subset of school Ofsted performance measures for which it is possible to obtain grades over the period between 2001 and 2008. No data on sub-grades is available before 2001.

and pupil achievement appear to be the only elements having a positive impact on salaries, while the change in the overall and the quality of management and leadership have coefficients which are negligible in size and statistically insignificant. These results suggest that the impact of ratings on principals' salary is felt via the evaluation of the quality of teaching that is being delivered in the school as well as through the evaluation of student achievement.

However, the pattern is very different if we look at the impact on the exit probability. Results for this outcome are reported in Panel (b) of Table 5. In this case, the quality of the management rating has a significant negative impact on the probability of exit when it is considered alone in Column (2). A grade increase in the rating for management and leadership reduces the probability of exit by 4.3 percentage point. Teaching quality and pupil achievement, whose impact is reported in Column (3) and Column (4) respectively, also have a negative impact on the probability of about 5 percentage points. The role of management and leadership becomes even starker when all measures are included jointly in the difference-in-differences in Column (5). In this case the impact of a change in management and leadership is lower in magnitude (-2.7 percentage points) but still sizeable and significant at 5%, while pupil achievement has a slightly larger impact (-2.9 percentage points) and it is significant at 10%. Teaching quality has the smallest impact (-1.9 percentage points) and it is statistically insignificant. Strikingly, the change in overall score has a negligible effect on the probability of exit. Similar results are obtained when all sub-scores are jointly included without the overall score in Column (6). Thus, there is a direct path between the evaluation of leadership and management and the retention outcome, as well as indirect link via the evaluation of student achievement.

This evidence suggests that direct rating of the quality of the management and leadership have important consequence for the retention decisions for the principals. In summary, while the overall inspection ratings are important for the principal's career and pay, more precise and granular assessment of their performance appears to be even more important. Furthermore, the analysis of the different components of the inspection ratings reveals that ratings for specific dimensions of school quality have an economically

meaningful impact on the career of principals and that the effect of the overall score is absorbed by these components.

7 Senior Leaders and Teachers

School inspections do not only have an impact on principals, but may also affect other school employees' pay and career progress. We extend the analysis to also examine the impact on senior leaders (deputy and assistant principals) and teachers. We expect that the more of a role a teacher plays in the governance of the school, the larger will be the impact on their pay or career progress. We consider the most relevant outcomes from previous analysis: salary, turnover, and school exit. We also add potential career progression measures such as: the probability of becoming a principal or a senior leader. Results are reported in Table 6 and are in line with our expectations: both senior leaders and teachers are affected by the change in inspection ratings, with the former experiencing the largest impact.

We also assess dynamics in some greater detail. We start by looking at senior leaders in Figure 5. As figure shows, they experience higher salaries by up to 1% after three years of inspection, as shown in Panel (a). In this case, a small difference in trends in log salary appears, especially in period $\tau = -3$. However, years closer to the inspection ($\tau = 0, -1, -2$), which matter most, do not show any suspicious pattern, with a very flat difference in wages. A similar pattern is observed for salary in levels in Panel (b). In order to study promotions, we classify teachers based on our broad categories in each year of our time span and assess whether inspection grade changes have an impact on the probability of teachers climbing the job ladder. We look at two main outcomes: becoming a senior leader (deputy or assistant head), in Panel (c), and becoming a principal, in Panel (d). Note that for senior leaders, becoming a principal is a promotion. Our findings show that a grade improvement increases the chances of senior leaders remaining in that role (but not in a statistically significant way) and of becoming principals. The effect is up to 1.5 percentage points. As these transitions are relatively

rare - for senior leaders only 6.8% of them become principals, on average, within three years of the inspection, relative to senior leaders in schools experiencing no change in rating - the effect of one grade improvement represents about 22% increase in the probability of becoming a principal. Overall, experiencing an improvement in Ofsted rating is beneficial for the career of both principals, who are more likely to keep their position, and senior leaders, who are more likely to be promoted.

Next, we assess the impact of these school rating changes on turnover. This is a particularly important margin as salary effects can be computed only for teachers who remain in the school sector and previous literature has shown that grade changes in Ofsted evaluation lead to higher turnover. To do so, we check two dimensions of mobility; first, whether the individual remains in the same school in which she was at the time of the inspection (Panel e); second, if they remain in the state-run school sector or not in (Panel f). We find an impact in line with the effect for principals, albeit, smaller in magnitude, as expected. When exiting schools, teachers could face two situations, which however, are very different for their career: first, they could move to another school, while remaining in the same sector; second, they could exit the school sector and either retire or change jobs. Panel (f) explores the latter case, by considering a dummy equal to one if the individual is no longer observed in any school. As can be seen, for senior leaders the consequence of exiting are quite dramatic, similarly to principals: they have a 1.8 percentage point lower probability of leaving their school per Ofsted grade increase and about 1.5 percentage point lower probability of exiting the school sector and retiring. The impact on principals was anyway substantially larger with almost a 6 percentage point decline in the probability of leaving the school and the school sector and a 3 percentage point decline in the probability of retirement.

Finally, we explore the impact on teachers in Figure 6. We again observe a positive impact on salaries, although smaller than that for school managers: teachers experiencing an improvement in their schools' Ofsted rating see an increase in their salary by the third year after the inspection by 0.2 percentage points or about £120 (the corresponding amount for principals was about 1.7%). Teachers reap benefits also in terms of career progression.

Panel (c) shows some transitory benefits for the probability of teachers becoming senior leaders: estimates show a temporary positive effect, up to 0.35 percentage point increase (over a baseline of 2.6% at $\tau = 2$), but this effect is not very stable and declines to become statistically insignificant in the last period. However, they experience an increase in their probability of becoming principals by nearly 0.1 percentage point as shown in Panel (d). Since the percentage of teachers becoming principals is only 0.27%, the effect represents more than a 33% increase in the baseline probability (although this effect is less precisely estimated). Looking at turnover, things are rather different for teachers: by the third year after the inspection they are about 1.5 percentage points more likely to remain in the same school, as shown in Panel (e), but only 0.5 percentage points less likely to have exited the profession, as reported in Panel (f). This suggests that while the managers face dramatic consequences for changes in their schools' Ofsted ratings, implications for teachers are much more modest. (Note that a little deviation from the parallel trend is observed in $\tau = -3$, but the magnitude of the coefficient is negligible.)

Throughout this analysis, the parallel trend assumption is supported by trend in years prior to the inspection in most of the cases, with only few exceptions. Panel (a) of Figure 5, that is the impact of a Ofsted rating change on salaries for senior leaders, in particular, shows some deviation from the parallel trend three years prior to the inspection. To assess to what extent the analysis is affected by this deviation we also present results for both senior leaders and teachers after excluding the 2007 cohort, which appears to be the most problematic in Figure A13 and Figure A12 in the Appendix. Both Figures, reassuringly, provide very similar results.

Overall, our results show that Ofsted ratings have an important impact on the careers of top level managers. The market for senior managers reacts strongly to external assessments, perhaps because they carry the greatest responsibility for the inspection outcome, while substantially smaller effects are observed for lower level employees.

8 Conclusion

In this paper we investigate the labor market impact of subjective school evaluations administered by external evaluators. We use data on English state-run secondary schools and the universe of their personnel matched with inspections undertaken between 2006 and 2008. This period was characterized by dramatic changes in the notice period provided to schools regarding the timing of the inspection visit which generated unexpected shocks to grades. We focus mainly on principals and compare, over time, principals in schools experiencing inspection grade changes (treated) with those in schools not experiencing any changes (comparison group) in a difference-in-differences framework.

After showing that treated and control schools and individuals exhibit similar characteristics and trends before the shock, we provide evidence that changes in inspection grades lead to 3% increase in their salary. However, these relatively modest wage effects mask large impacts along the extensive margin. A change in rating leads to a 6 percentage point decline in the probability of principals exiting the (public sector) teacher labor market. This overall impact of a change in rating is highly asymmetric, driven largely by declines in ratings. Thus, the assessments would appear to carry large downside risks for principals, with much more modest gains on the upside. Finally, changes in inspection grades also affect lower level managers and teachers, but to a much smaller degree relative to principals.

At a theoretical level, there are a number of explanations behind the dramatic effects we find for school principals. If a school's improved inspection rating provides a signal of the underlying productivity of the principal, then this may lead the wider market, including local school boards engaged in hiring a principal, to update their priors, resulting in greater bargaining power for the incumbent principal. A somewhat different explanation hinges on a performance pay channel. In particular, if ratings induce greater parental demand and impact the socioeconomic composition of students, ratings may be valued in and of themselves by the school's governing board. Hence, if principals are

rewarded for delivering better inspection outcomes, there may be an important role for this performance pay mechanism.

We exploit a unique feature of our setting where both the overall organizational quality rating as well as other margins of quality, in particular, ratings for the quality and effectiveness of the leadership and management, are available in order to test the relevance of each of these channels. We find evidence suggesting that both channels are at play. In particular, our results demonstrate that there is a direct and powerful link between the evaluation of leadership and management and the retention-exit margin for principals. This points to an information channel. We also uncover evidence that improvements in ratings lead to higher levels of school enrolment and a fall in the proportion of students eligible for free lunch, over and above any impact arising from any other school quality measures. This would suggest that school boards may indeed value higher ratings because they induce these demand side responses.

These results provide novel evidence on the impact of subjective quality assessments of public services, in particular in the education sector. Our findings demonstrate that quality assessments provide additional and new signals to schools' boards and the market, generating significant consequences for school managers' careers.

References

Besley, T., Machin, S., and Telhaj, S. (2021). Are Public CEOs Different? Pay for Performance and Salary Inequality. *mimeo*.

Borusyak, K., Jaravel, X., and Spiess, J. (2021). Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*.

Burgess, S., Greaves, E., and Murphy, R. (2022). Deregulating teacher labor markets. *Economics of Education Review*, 88:102253.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with Multiple Time Periods. *Journal of Econometrics*, 225(2):200–230.

Carnoy, M. and Loeb, S. (2002). Does external accountability affect student outcomes? a cross-state analysis. *Educational evaluation and policy analysis*, 24(4):305–331.

Chevalier, J. and Ellison, G. (1999). Career concerns of mutual fund managers. *The Quarterly Journal of Economics*, 114(2):389–432.

Cullen, J. B., Hanushek, E. A., Phelan, G., and Rivkin, S. G. (2021). Performance information and personnel decisions in the public sector: The case of school principals. *Journal of Human Resources*, pages 0619–10272R1.

De Janvry, A., He, G., Sadoulet, E., Wang, S., and Zhang, Q. (2023). Subjective performance evaluation, influence activities, and bureaucratic work behavior: Evidence from china. *American Economic Review*, 113(3):766–799.

Dee, T. S., Jacob, B., and Schwartz, N. L. (2013). The effects of nclb on school resources and practices. *Educational Evaluation and Policy Analysis*, 35(2):252–279.

Dee, T. S., Jacob, B. A., Hoxby, C. M., and Ladd, H. F. (2010). The impact of no child left behind on students, teachers, and schools [with comments and discussion]. *Brookings papers on economic activity*, pages 149–207.

Deming, D. J., Cohodes, S., Jennings, J., and Jencks, C. (2016). School accountability, postsecondary attainment, and earnings. *Review of Economics and Statistics*, 98(5):848–862.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of human resources*, pages 696–727.

Dizon-Ross, R. (2020). How Does School Accountability Affect Teachers? Evidence from New York City. *Journal of Human Resources*, 55(1):76–118.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of economic literature*, 48(4):935–63.

Dranove, D., Kessler, D., McClellan, M., and Satterthwaite, M. (2003). Is more information better? the effects of "report cards" on health care providers. *Journal of political Economy*, 111(3):555–588.

Feng, L., Figlio, D., and Sass, T. (2018). School Accountability and Teacher Mobility. *Journal of Urban Economics*, 103:1–17.

Figlio, D. N. and Winicki, J. (2005). Food for Thought: the Effects of School Accountability Plans on School Nutrition. *Journal of Public Economics*, 89(2-3):381–394.

Frederiksen, A., Lange, F., and Kriechel, B. (2017). Subjective Performance Evaluations and Employee Careers. *Journal of Economic Behavior & Organization*, 134:408–429.

Gibbons, R. and Katz, L. F. (1991). Layoffs and lemons. *Journal of labor Economics*, 9(4):351–380.

Gjefsen, H. M. and Gunnes, T. (2020). How School Accountability Affects Teacher Mobility and Teacher Sorting. *Education Economics*, 28(5):455–473.

Greaves, E., Hussain, I., Rabe, B., and Rasul, I. (2023). Parental Responses to Information about School Quality: Evidence from Linked Survey and Administrative Data. *The Economic Journal*.

Greene, J. P. et al. (2001). *An evaluation of the Florida A-Plus accountability and school choice program*. Center for Civic Innovation at the Manhattan Institute.

Hanushek, E. A. and Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 24(2):297–327.

Hayes, R. M. and Schaefer, S. (2000). Implicit contracts and the explanatory power of top executive compensation for future performance. *The RAND Journal of Economics*, pages 273–293.

Hernandez-Arenaz, I. and Iriberri, N. (2019). A review of gender differences in negotiation. *Oxford Research Encyclopedia of Economics and Finance*.

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, 7:24.

Hussain, I. (2015). Subjective Performance Evaluation in the Public Sector Evidence from School Inspections. *Journal of Human Resources*, 50(1):189–221.

Hussain, I. (2022). Housing Market and School Choice Response to School Quality Information Shocks. *mimeo*.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the chicago public schools. *Journal of public Economics*, 89(5-6):761–796.

Ladd, H. F. (1999). The Dallas School Accountability and Incentive Program: An Evaluation of its Impacts on Student Outcomes. *Economics of Education Review*, 18(1):1–16.

Lazear, E. P. and Oyer, P. (2012). Personnel Economics. In Gibbons, R. and Roberts, J., editors, *The Handbook of Organizational Economics*, pages 479–520. Princeton University Press, Princeton, NJ.

Li, D. (2015). School Accountability and Principal Mobility: How No Child Left Behind affects the allocation of school leaders. *Harvard Business School Working Paper N. 16-052*.

Loeb, S. and Cunha, J. (2007). Have assessment-based accountability reforms influenced the career decisions of teachers and principals? *Report commissioned by US Congress, Prepared as part of Title I, Part E, Section*, 1503.

Luginbuhl, R., Webbink, D., and De Wolf, I. (2009). Do inspections improve primary school performance? *Educational evaluation and policy analysis*, 31(3):221–237.

Machin, S., McNally, S., and Ruiz-Valenzuela, J. (2020). Entry through the Narrow Door: the Costs of Just Failing High Stakes Exams. *Journal of Public Economics*, 190:104224.

Mansfield, J. and Slichter, D. (2021). The Long-Run Effects of Consequential School Accountability. *IZA Discussion Paper n. 14.503*.

Milgrom, P. and Roberts, J. (1988). An Economic Approach to Influence Activities in Organizations. *American Journal of Sociology*, 94:S154–S179.

Murphy, K. J. (1999). Executive compensation. *Handbook of labor economics*, 3:2485–2563.

Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of economic literature*, 37(1):7–63.

Prendergast, C. and Topel, R. H. (1996). Favoritism in Organizations. *Journal of Political Economy*, 104(5):958–978.

Richards, C. E. and Sheu, T. M. (1992). The south carolina school incentive reward program: A policy analysis. *Economics of Education review*, 11(1):71–86.

Scanlon, M. (1999). *The Impact of Ofsted Inspections*. National Foundation for Education Research for the National Union of Teachers, Slough.

Schervish, M. J. and DeGroot, M. H. (2012). *Probability and statistics*. Pearson Education.

Schönberg, U. (2007). Testing for asymmetric employer learning. *Journal of Labor Economics*, 25(4):651–691.

Tables

Table 1: Summary Statistics and Difference in Characteristics across Grade Changes.

Variable	(1) Baseline (No change)	(2) Worsening	(3) T-Stat Difference (2)-(1)	(4) Improvement	(5) T-Stat Difference (4)-(1)
			Schools		
N. pupils	1048.288	1076.626	1.634	994.403	-2.012
Pupil Teacher Ratio	17.202	16.628	-1.552	16.982	-0.488
Share FSM	0.001	0.001	0.896	0.002	2.463
Share No Pass	0.017	0.023	5.046	0.014	-1.243
Share 5+ A*-C GCSE	0.614	0.558	-6.344	0.628	1.09
Share Entry	0.173	0.178	1.065	0.182	1.333
Total Teachers	64.488	66.325	1.508	60.377	-2.298
Total Managers	5.109	5.139	0.246	5.23	0.578
Female	41.759	43.041	1.527	39.416	-1.99
			Principals		
Salary Head	72723.343	71521.125	-1.957	73699.21	0.952
Female Haed	0.344	0.312	-1.261	0.33	-0.365
Age Head	52.057	51.588	-1.781	51.478	-1.379
Tenure Head	6.427	5.69	-3.426	6.061	-1.182
N. Schools	742	729		209	
			Senior Leaders		
Salary	48665.894	48133.570	-3.513	48810.257	.595
Female	.437	.425	-.851	.433	-.119
Age	47.828	48.133	1.618	46.839	-3.299
Part time	.009	.012	.601	.008	-.426
Tenure	4.538	4.421	-1.460	4.170	-3.150
N. Managers	3188	3185		907	
			Teachers		
Salary	33835.527	33814.222	-.458	34025.660	2.552
Female	.597	.603	1.717	.606	1.495
Age	43.063	43.076	.159	42.889	-1.386
Part time	.143	.141	-.523	.127	-3.835
Tenure	7.092	7.185	2.815	6.908	-3.523
N. Teachers	33122	33362		8367	

Note: Table reports summary statistics for schools and teachers experiencing an inspection between 2006 and 2008 in the year before the inspection. The sample includes only schools with no grade change in the three years prior to the inspection and with starting grade above "Fail". School personnel is included in the sample if they were in the school for at least two years (the year preceding the inspection and the year of the inspection). Bottom line reports the number of schools by category of grade change. Column (1) reports the average for schools experiencing no change in their Ofsted grade, Column (2) reports the average for schools experiencing a grade decline, and Column (4) reports the average for schools experiencing an increase in their grade. Column (3) displays the t- statistic for the difference between Column (2) and Column (1), while Column (5) displays the t-statistic for the difference between Column (5) and Column (1). T-statistics based on robust standard errors.

Table 2: Effect of Grade Change on principals' Salaries

Variable	(1) Log Salary	(2) Log Salary	(3) Log Salary	(4) Log Salary	(5) Salary	(6) Salary (Miss)
Grade Change X Post	0.017***	0.016***	0.019***	0.018***	1,603.520***	5,232.254***
	(0.006)	(0.006)	(0.006)	(0.006)	(320.701)	(829.822)
Observations	10,926	10,926	10,926	10,926	10,926	11,753
Baseline					72516.136	
R-squared	0.213	0.722	0.731	0.737	0.869	0.548
Age Polynomial	YES	YES	YES	YES	YES	YES
IndividualXCohort FE	NO	YES	YES	YES	YES	YES
Year Fe	NO	YES	YES	YES	YES	YES
Cohort FE	NO	YES	YES	YES	YES	YES
Tenure FE	NO	NO	YES	YES	YES	YES
CohortXYear FE	NO	NO	YES	YES	YES	YES
TenureXCohort FE	NO	NO	YES	YES	YES	YES
TenureXYear FE	NO	NO	YES	YES	YES	YES
TenureXCohortXYear FE	NO	NO	NO	YES	YES	YES

Note: Table reports estimates of the effect of a change in Ofsted grade for principals based on Equation 1. Panel covers years ranging from three years before to three years after the inspection (inspections taking place between 2006 and 2008). Tenure is computed in the year before the inspection. Standard errors are clustered at the school level at the time of inspection. Level of Significance: *** 0.01, ** 0.05, * 0.1.

Table 3: Effect of Grade Change on Principals' Career: Principal's characteristics

VARIABLES	(1) Baseline	(2) Male	(3) Female	(4) Older	(5) Younger	(6) High Tenure	(7) Low Tenure
Panel (a): (log) Salary							
Grade Change X Post	0.018***	0.023***	0.006	0.007	0.028***	0.020***	0.013
	(0.006)	(0.007)	(0.009)	(0.008)	(0.008)	(0.007)	(0.009)
Observations	11,137	7,445	3,684	6,015	5,099	8,452	2,685
R-squared	0.727	0.701	0.824	0.749	0.708	0.693	0.801
Panel (b): Exit							
Grade Change X Post	-0.044***	-0.045***	-0.033**	-0.066***	-0.025***	-0.053***	-0.021
	(0.009)	(0.011)	(0.015)	(0.014)	(0.009)	(0.011)	(0.014)
Observations	11,977	8,022	3,948	6,741	5,215	9,212	2,765
R-squared	0.499	0.517	0.507	0.526	0.365	0.516	0.320
Age Polynomial	YES	YES	YES	YES	YES	YES	YES
IndividualXCohort FE	YES	YES	YES	YES	YES	YES	YES
Year Fe	YES	YES	YES	YES	YES	YES	YES
Cohort FE	YES	YES	YES	YES	YES	YES	YES
Tenure FE	YES	YES	YES	YES	YES	YES	YES
CohortXYear FE	YES	YES	YES	YES	YES	YES	YES
TenureXCohort FE	YES	YES	YES	YES	YES	YES	YES
TenureXYear FE	YES	YES	YES	YES	YES	YES	YES
TenureXCohortXYear FE	YES	YES	YES	YES	YES	YES	YES

Note: Table reports estimates of the effect of a change in Ofsted grade for principals based on Equation 1. Panel covers years ranging from three years before to three years after the inspection (inspections taking place between 2006 and 2008). Tenure is computed in the year before the inspection. Standard errors are clustered at the school level at the time of inspection. Level of Significance: *** 0.01, ** 0.05, * 0.1.

Table 4: Effect of Grade Change on principals' Career: School Characteristics and Competition

VARIABLES	(1) Baseline	(2) School Char	(3) n.schools $< 50p$	(4) n.schools $> 50p$
Grade Change X Post	0.018***	0.016***	0.020**	0.014*
	(0.006)	(0.006)	(0.008)	(0.009)
Observations	11,137	10,874	5,928	5,209
R-squared	0.727	0.740	0.698	0.777
Panel (b): Exit				
Grade Change X Post	-0.044***		-0.046***	-0.041***
	(0.009)		(0.012)	(0.013)
Observations	11,977	10,874	6,405	5,572
R-squared	0.499		0.522	0.502
Age Polynomial	YES	YES	YES	YES
IndividualXCohort FE	YES	YES	YES	YES
Year Fe	YES	YES	YES	YES
Cohort FE	YES	YES	YES	YES
Tenure FE	YES	YES	YES	YES
CohortXYear FE	YES	YES	YES	YES
TenureXCohort FE	YES	YES	YES	YES
TenureXYear FE	YES	YES	YES	YES
TenureXCohortXYear FE	YES	YES	YES	YES

Note: Table reports estimates of the effect of a change in Ofsted grade for principals based on Equation 1. Panel covers years ranging from three years before to three years after the inspection (inspections taking place between 2006 and 2008). Tenure is computed in the year before the inspection. Column (1) reports the baseline estimates for the sake of comparison. Column (2) includes controls for time varying characteristics of the schools such as (log) number of students, share of students on Free School Meal, Share of students with Special Education Needs, Share of students achieving no passes, and share of students with 5 or more A* to C in GCSE exam; Column (3) reports the impact of a change in Ofsted grade when there are less than the median number of schools within a 5 miles radius from the school; Finally, Column (4) reports the impact of a change in Ofsted grade when there are more than the median number of schools within a 5 miles radius from the school. Standard errors are clustered at the school level at the time of inspection. Level of Significance: *** 0.01, ** 0.05, * 0.1.

Table 5: Effect of Grade Change on principals' Salaries and Career: Subscores

VARIABLES	(1) Salary	(2) Leadership	(3) Teaching	(4) Achievement	(5) All subscores	(6) All subscores
			Panel (a): (log) Salaries			
Grade Change X Post	0.017**				-0.011	
	(0.007)				(0.013)	
Management and Leadership X Post		0.011**			-0.003	-0.006
		(0.005)			(0.007)	(0.007)
Teaching Quality X Post			0.028***		0.020*	0.017*
			(0.006)		(0.011)	(0.010)
Pupil Achievement X Post				0.026***	0.025**	0.020*
				(0.006)	(0.011)	(0.011)
Observations	8,517	8,517	8,510	8,510	8,503	8,503
Age Polynomial	YES	YES	YES	YES	YES	YES
F-test Jointly diff. 0					0	0
F-test Jointly diff.					.037	.033
			Panel (b): Exit			
Grade Change X Post	-0.043***				0.010	
	(0.010)				(0.018)	
Management and Leadership X Post		-0.043***			-0.027**	-0.024**
		(0.009)			(0.012)	(0.011)
Teaching Quality X Post			-0.051***		-0.019	-0.016
			(0.012)		(0.017)	(0.017)
Pupil Achievement X Post				-0.049***	-0.029*	-0.024*
				(0.010)	(0.017)	(0.014)
Observations	9,170	9,170	9,163	9,163	9,156	9,156
R-squared	0.506	0.506	0.506	0.507	0.508	0.508
F-test Jointly diff. 0					0	0
F-test Jointly diff.					.54	.937
IndividualXCohort FE	YES	YES	YES	YES	YES	YES
Year Fe	YES	YES	YES	YES	YES	YES
Cohort FE	YES	YES	YES	YES	YES	YES
Tenure FE	YES	YES	YES	YES	YES	YES
CohortXYear FE	YES	YES	YES	YES	YES	YES
TenureXCohort FE	YES	YES	YES	YES	YES	YES
TenureXYear FE	YES	YES	YES	YES	YES	YES
TenureXCohortXYear FE	YES	YES	YES	YES	YES	YES

Note: Table reports estimates of the effect of a change in Ofsted grade for principals based on Equation 1. Panel covers years ranging from three years before to three years after the inspection (inspections taking place between 2006 and 2008). management and Leadership, Teaching Quality and Pupil Achievement range from 1 to 4 and are subscore for the Ofsted for specific aspects of the performance of the school. The table includes only inspections for which a change in each subscore could be computed and Column (1) reports estimates for our baseline model including only observations with available subscore changes (about half of the inspections of 2006 are lost due to absent data on subscores for the precious inspection). Tenure is computed in the year before the inspection. Standard errors are clustered at the school level at the time of inspection. Level of Significance: *** 0.01, ** 0.05, * 0.1.

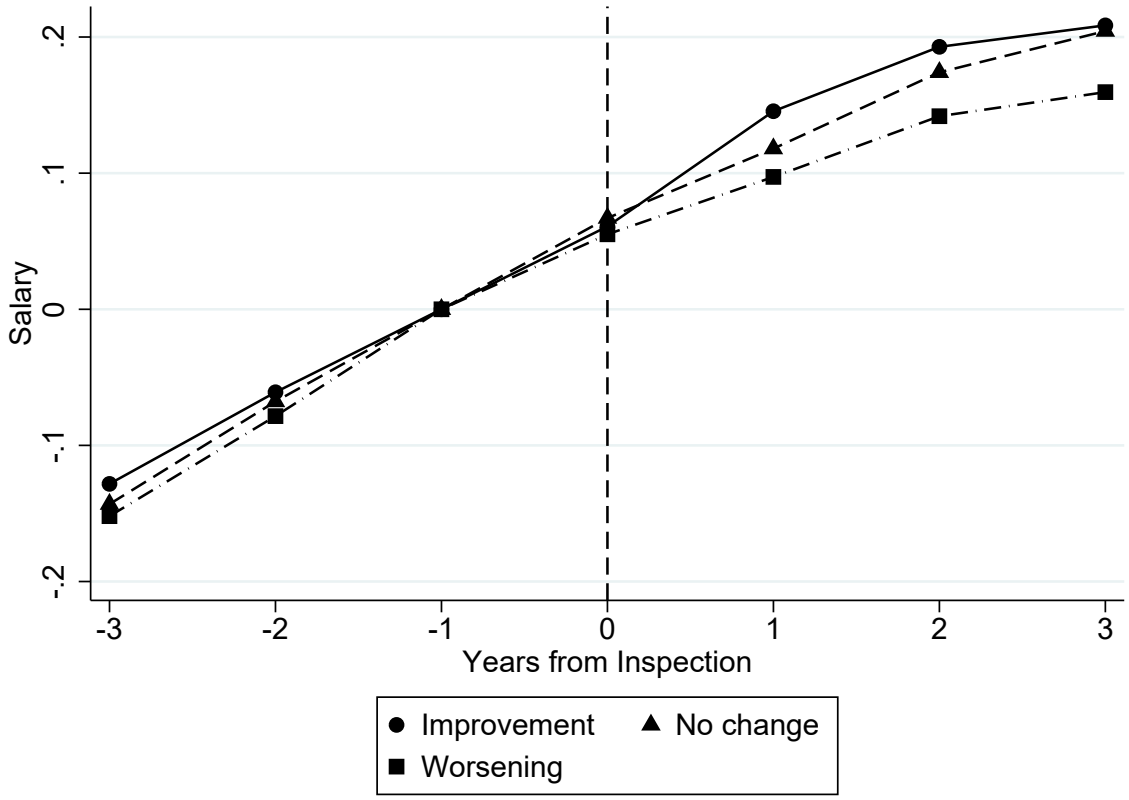
Table 6: Effect of Grade Change on Senior Leaders and Teachers' Salaries

Variables	(1) Log Salary	(2) Salary	(3) Manager	(4) Principal	(5) Leave School	(6) Exit
			Panel (a): Senior Leaders			
Grade Change X Post	0.009***	530.358***	0.010	0.007**	-0.009*	-0.009**
	(0.002)	(104.036)	(0.006)	(0.003)	(0.005)	(0.004)
Baseline		55662.661	0.804	.05	.18	.095
Observations	48,652	48,657	50,960	50,960	50,960	50,960
			Panel (b): Teachers			
Grade Change X Post	0.002***	100.876***	0.002**	0.000**	-0.011***	-0.004**
	(0.001)	(25.828)	(0.001)	(0.000)	(0.003)	(0.002)
Baseline		38611.706	.026	.003	.221	.12
Observations	493,470	493,476	523,957	523,957	523,957	523,957
Age Polynomial	YES	YES	YES	YES	YES	YES
Individual FE	YES	YES	YES	YES	YES	YES
Year Fe	YES	YES	YES	YES	YES	YES
Cohort FE	YES	YES	YES	YES	YES	YES
Tenure FE	YES	YES	YES	YES	YES	YES
CohortXYear FE	YES	YES	YES	YES	YES	YES
TenureXCohort FE	YES	YES	YES	YES	YES	YES
TenureXYear FE	YES	YES	YES	YES	YES	YES
TenureXCohortXYear FE	YES	YES	YES	YES	YES	YES

Note: Table reports estimates for the effect of a change in Ofsted grade for principals based on Equation 1. Panel covers years ranging from three years before to three years after the inspection (we consider only inspections taking place between 2006 and 2008 for reasons outlined in the paper). The coefficient of the effect of the inspection on the probability of becoming principal for teachers (Column 4, Panel b) is .00046. Tenure is computed in the year before the inspection. Standard errors are clustered at the school level at the time of inspection. Level of Significance: *** 0.01, ** 0.05, * 0.1.

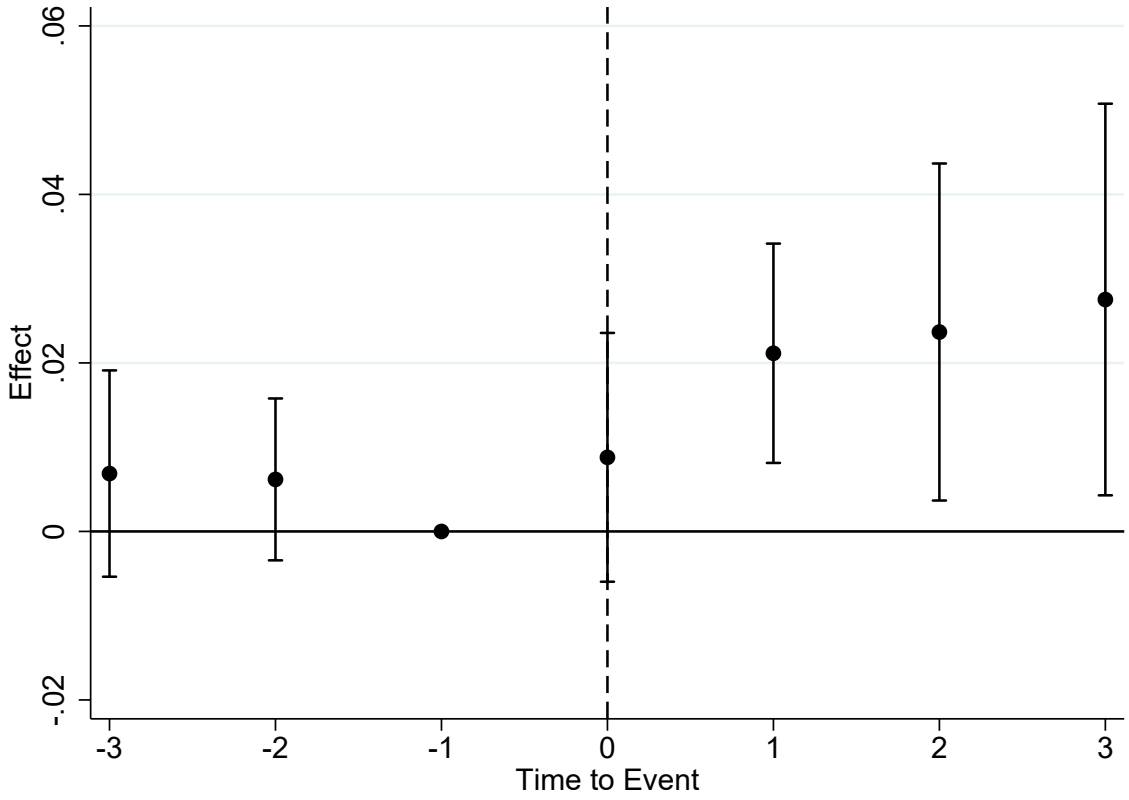
Figures

Figure 1: Mean (log) Salary for Principals around Inspections by Grade Change



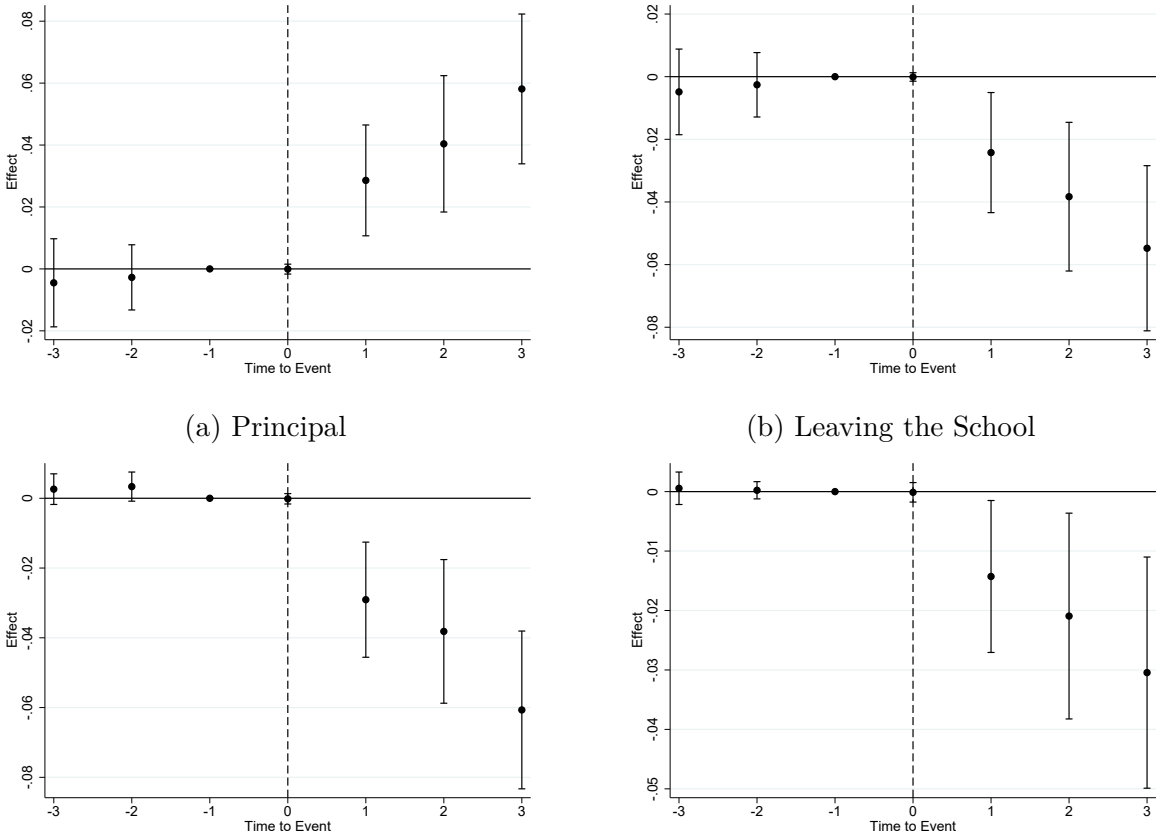
Notes: Average (log) salaries for principals normalized at the year before the inspection ($\tau = -1$) for principals experiencing an inspection between 2006 and 2008. Salaries are reported separately for improvement, decline and stable inspection rating.

Figure 2: Effect of Change in Ofsted Grade on principals' (log) Salaries



Notes: Figure reports the effect of Change in Ofsted grade on log salaries based on 2 as well as 95% confidence intervals. Regression includes an age polynomial and fixed effects at individual-cohort level and at the year-cohort-tenure level. The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

Figure 3: Effect of Change in Ofsted Grade on principals' Career and Turnover



(a) Principal

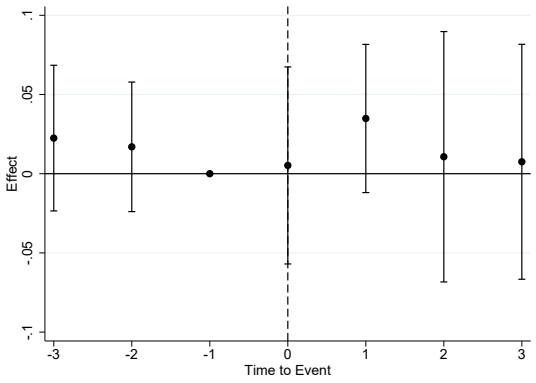
(b) Leaving the School

(c) Exit

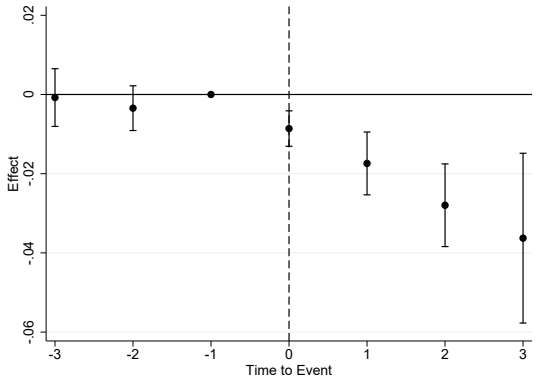
(d) Pension

Notes: Figure reports the effect of Change in Ofsted grade on career outcomes based on Equation 2 as well as 95% confidence intervals. Panel (a) reports the effect on the probability of being a principal; Panel (b) reports the effect on the probability of leaving the school; Panel (c) reports the effects on the probability of exiting the school sector; and Panel (d), finally, reports the effect on the probability of retirement. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. The baseline period is the year before the inspection $\tau = -1$. Standard errors are clustered at the school level at the time of inspection.

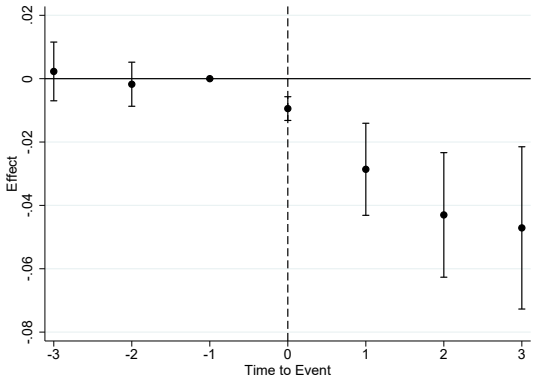
Figure 4: Effect of Change in Ofsted Grade on principals' (log) Salaries: Improvement and Decline



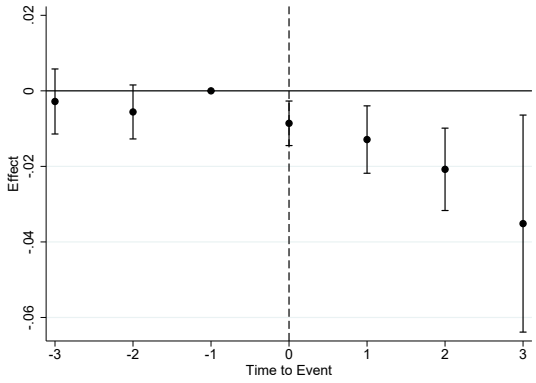
(a) Improvement



(b) Worsening



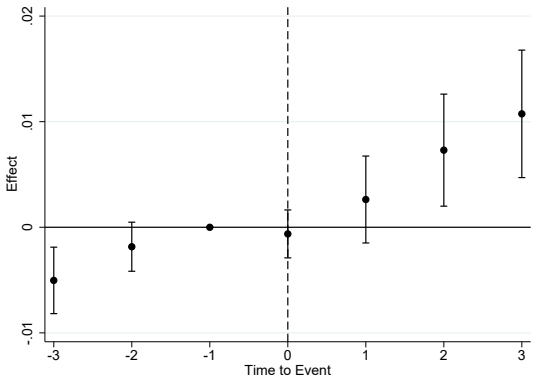
(c) Worsening - Fail



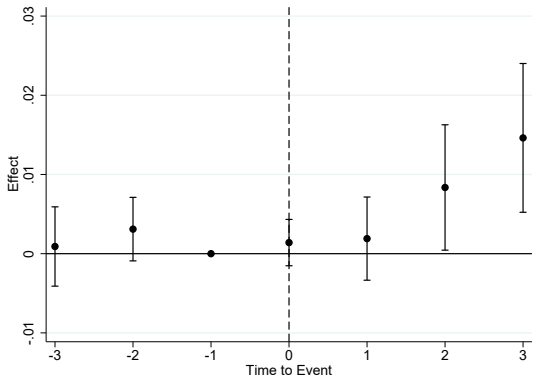
(d) Worsening - No Fail

Notes: Figure reports the effect of Change in Ofsted grade on log salaries based on Equation 2 as well as 95% confidence intervals. Panel (a) reports the effect of a grade improvement; Panel (b) reports the effect of a grade worsening; Panel (c) reports the effects of a grade worsening which does not lead to a "fail" grade for the school; and Panel (d), finally, reports the effect of a grade worsening leading to a fail grade. The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

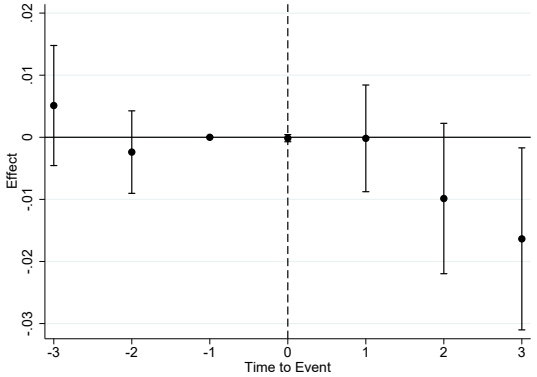
Figure 5: Effect of Change in Ofsted Grade on Senior Leaders' Salary and Career



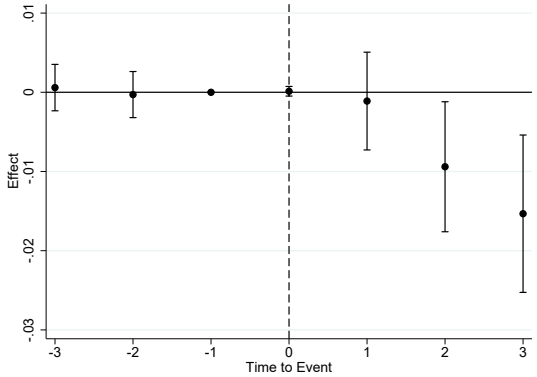
(a) (Log) Salary



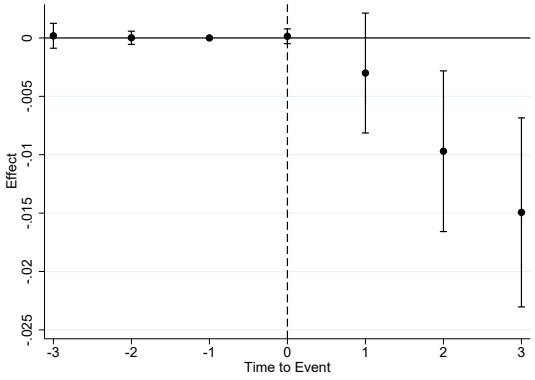
(b) Principal



(c) Leave School



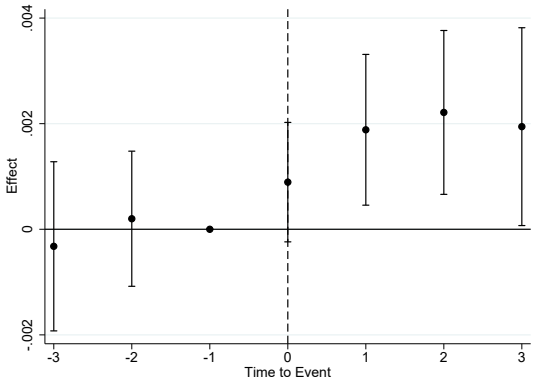
(d) Exit



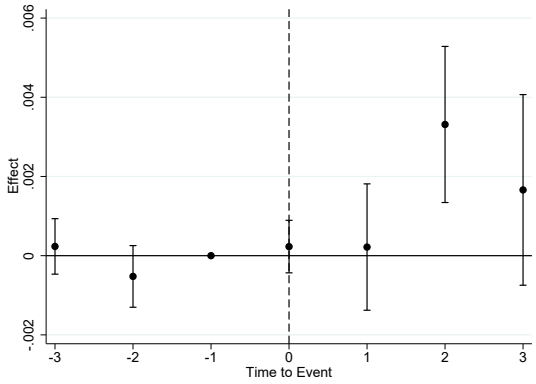
(e) Pension

Notes: Figure reports the effect of Change in Ofsted grade on Senior Leaders' outcomes based on Equation 2 as well as 95% confidence intervals. Panel (a) reports the effect on log salary; Panel (b) reports the effects on the probability of becoming a principal; Panel (c) reports the effects on the probability of changing school; Panel (d) reports the effect on the probability of exiting the school sector; and Panel (e) reports the effect on the probability of being in retirement. The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at the time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

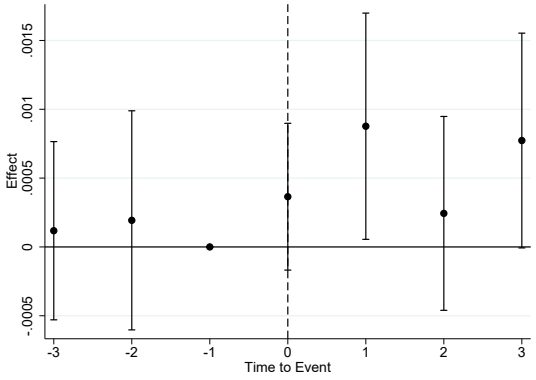
Figure 6: Effect of Change in Ofsted Grade on Teachers' Salary and Career



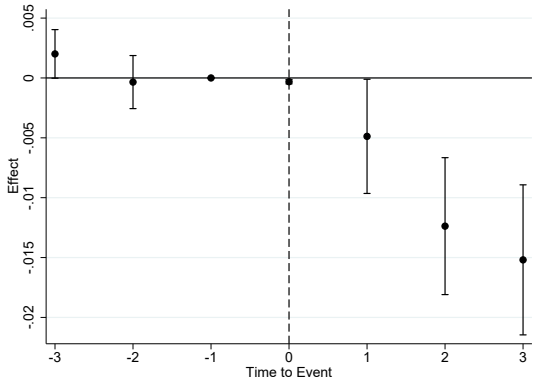
(a) (Log) Salary



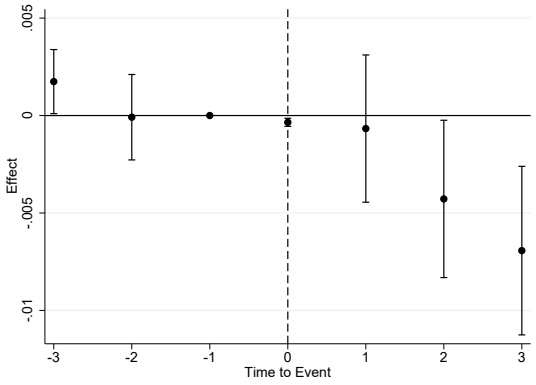
(b) Senior Leader



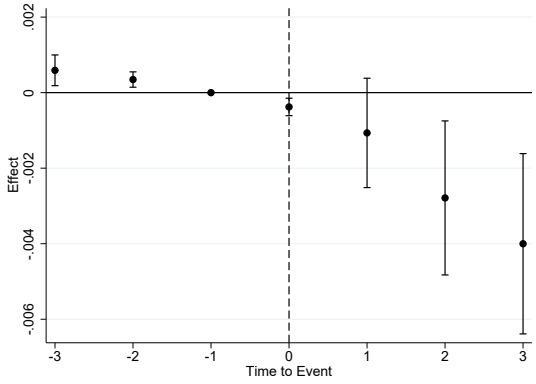
(c) Principal



(d) Leave School



(e) Exit



(f) Pension

Notes: Figure reports the effect of Change in Ofsted grade on teachers' outcomes based on Equation 2 as well as 95% confidence interval. Panel (a) reports the effect on log salary; Panel (b) reports the effects on the probability of being a senior leader; Panel (c) reports the effects on the probability of becoming a principal; Panel (d) reports the effects on the probability of changing school; Panel (e) reports the effect on the probability of exiting the school sector and Panel (e) reports the effect on the probability of being in retirement. The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

Appendix

Tables

Table A1: Baseline and Changes in Ofsted grades between 2006 and 2008

	(1)	(2)	(3)	(4)
		Previous Grade		
Change Ofsted Grade	2 (Requires Improvement)	3 (Good)	4 (Outstanding)	Total by Grade Change
-3			12	12
-2		58	60	118
-1	57	331	211	599
0	227	393	121	741
1	144	52		196
2	13			13
Total	441	834	404	1679

Note: Table reports the number of schools by change in Ofsted grade and baseline rating in our sample. Sample restrictions described in Section 3.3.

Table A2: Tabulation of Distance in years between inspections

Distance	# Schools 2006	Share 2006	# Schools 2007	Share 2007	# Schools 2008	Share 2008
1	0	0.00	41	0.04	9	0.01
2	25	0.03	25	0.03	5	0.01
3	5	0.01	41	0.04	215	0.31
4	20	0.02	281	0.31	332	0.48
5	319	0.39	312	0.34	68	0.10
6	404	0.49	158	0.17	25	0.04
7	4	0.00	23	0.03	7	0.01
8	0	0.00	7	0.01	7	0.01
9	3	0.00	4	0.00	18	0.03
10	8	0.01	17	0.02	2	0.00
11	27	0.03	6	0.01	2	0.00
12	3	0.00	1	0.00	0	0.00
13	0	0.00	0	0.00	0	0.00
14	0	0.00	0	0.00	1	0.00

Note: Tabulation of the distance in years between inspections for school receiving an inspection between 2006 and 2008.

Table A3: Observable characteristics and probability of inspection.

VARIABLES	(1) Inspection	(2) Inspection	(3) Inspection	(4) Inspection	(5) Inspection
Distance past Insp: 2 years	0.004	0.022***	0.031***	0.031***	0.032***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Distance past Insp: 3 years	0.155***	0.192***	0.203***	0.204***	0.209***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Distance past Insp: 4 years	0.405***	0.430***	0.443***	0.443***	0.449***
	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)
Distance past Insp: 5 years	0.555***	0.579***	0.602***	0.602***	0.603***
	(0.015)	(0.014)	(0.015)	(0.015)	(0.015)
Distance past Insp: 6 years	0.672***	0.697***	0.726***	0.727***	0.726***
	(0.017)	(0.016)	(0.017)	(0.017)	(0.017)
Distance past Insp: 7 years or more	0.359***	0.373***	0.384***	0.384***	0.362***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Past Ofsted: Good		-0.115***	-0.111***	-0.112***	-0.130***
		(0.009)	(0.009)	(0.009)	(0.009)
Past Ofsted: Outstanding		-0.266***	-0.260***	-0.261***	-0.296***
		(0.010)	(0.010)	(0.011)	(0.013)
Year 2007			0.096***	0.096***	0.090***
			(0.009)	(0.009)	(0.009)
Year 2008			0.091***	0.091***	0.076***
			(0.010)	(0.010)	(0.010)
Female (head)				-0.006	0.001
				(0.008)	(0.008)
Age (head)				-0.000	0.000
				(0.001)	(0.001)
New Head (head)				-0.010	-0.005
				(0.009)	(0.009)
Share of FSM eligible					-0.001***
					(0.000)
Share no passes (GCSE)					-0.005**
					(0.002)
Share 5+ A*-C (GCSE)					0.001*
					(0.000)
log N. Pupils					0.005
					(0.011)
Change (2 years) Share of FSM eligible					0.002*
					(0.001)
Change (2 years) Share no passes (GCSE)					0.001
					(0.002)
Change (2 years) Share 5+ A*-C (GCSE)					0.001
					(0.000)
Change (2 years) log N. Pupils					0.010
					(0.053)
Observations	8,697	8,697	8,697	8,690	8,522
R-squared	0.312	0.359	0.368	0.369	0.374

Note: Linear probability model with dependent variable a dummy equal to one if the the school is subject to an inspection in the year. Model includes dummies for the distance from the last Ofsted inspection, past inspection grade, year dummies, characteristics of the principal (gender, age and if the school changed head in the year or in the previous year), and school characteristics (share of students with no passes in the GCSE, share of students with share of 5+ A*-C, log of the the number of students and share of students eligible to free school meal, FSM). Robust standard errors reported in parenthesis. Level of Significance: *** 0.01, ** 0.05, * 0.1.

Table A4: Effect of Unexpected Grade Change on principals' Salaries

VARIABLES	(1) Log Salary	(2) Log Salary	(3) Log Salary	(4) Log Salary	(5) Salary	(6) Salary (Miss)
Unexpected Grade Change X Post	0.020***	0.015***	0.015***	0.018***	1,951.542***	6,288.254***
	(0.006)	(0.005)	(0.005)	(0.005)	(347.615)	(1,024.123)
Observations	10,962	10,962	10,962	10,962	10,962	11,788
R-squared	0.209	0.709	0.709	0.724	0.863	0.542
Age Polynomial	YES	YES	YES	YES	YES	YES
IndividualXCohort FE	NO	YES	YES	YES	YES	YES
Year Fe	NO	YES	YES	YES	YES	YES
Cohort FE	NO	YES	YES	YES	YES	YES
Tenure FE	NO	NO	YES	YES	YES	YES
CohortXYear FE	NO	NO	NO	YES	YES	YES
TenureXCohort FE	NO	NO	NO	YES	YES	YES
TenureXYear FE	NO	NO	NO	YES	YES	YES
TenureXCohortXYear FE	NO	NO	NO	YES	YES	YES
Baseline					72389.372	

Note: Table reports estimates of the effect of an unexpected change in Ofsted grade for principals based on Equation 1. Panel covers years ranging from three years before to three years after the inspection (inspections taking place between 2006 and 2008). Tenure is computed in the year before the inspection. Unexpected grade change computed by using an ordered probit for the grade in the upcoming inspection with school characteristics in the year before the inspection, change in school characteristics between two years before the inspection and the year before the inspection and the past inspection grade. The school is then assigned the grade with the highest probability. Standard errors are clustered at the school level at the time of inspection. Level of Significance: *** 0.01, ** 0.05, * 0.1.

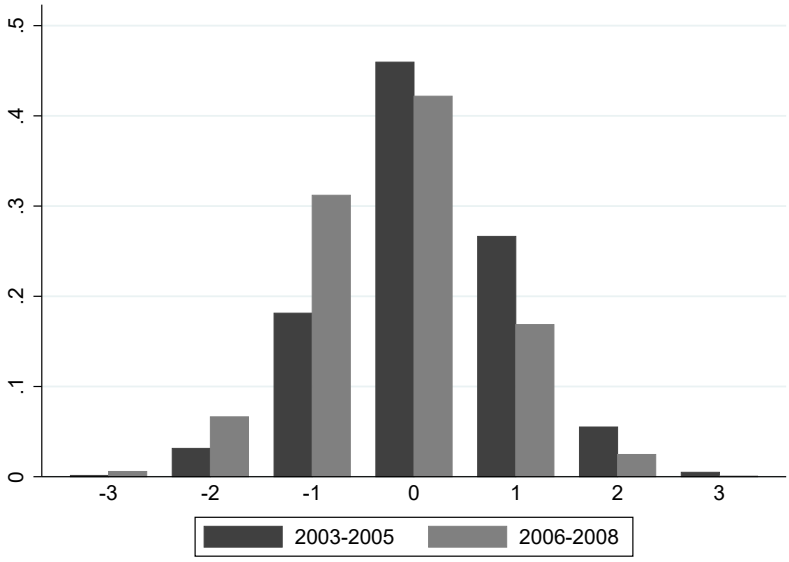
Table A5: Effect of Grade Change on School Personnels' Salaries and Career: Subscores

VARIABLES	(1) Subscore present	(2) Leadership	(3) Teaching	(4) Achievement	(5) Horse Race	(6) Horse Race
			Senior Leaders			
Grade Change X Post	0.009***				-0.000	
	(0.002)				(0.004)	
changeleadmanXpost		0.007***			0.002	0.002
		(0.002)			(0.003)	(0.002)
changeteachXpost			0.010***		0.003	0.003
			(0.002)		(0.003)	(0.003)
changepupacXpost				0.011***	0.007**	0.007**
				(0.002)	(0.003)	(0.004)
Observations	38,194	38,194	38,166	38,122	38,094	38,094
R-squared	0.815	0.815	0.815	0.815	0.816	0.816
F-test Jointly diff. 0					0	0
F-test Jointly diff.					.515	0.573
			Teachers			
Grade Change X Post	0.002**				0.001	
	(0.001)				(0.002)	
changeleadmanXpost		-0.000			-0.002**	-0.002**
		(0.001)			(0.001)	(0.001)
changeteachXpost			0.002**		0.002	0.002*
			(0.001)		(0.001)	(0.001)
changepupacXpost				0.002**	0.001	0.002
				(0.001)	(0.001)	(0.001)
Observations	380,693	380,693	380,376	380,413	380,096	380,096
F-test Jointly diff. 0					.017	.01
F-test Jointly diff.					.045	.019
Age Polynomial	YES	YES	YES	YES	YES	YES
IndividualXCohort FE	YES	YES	YES	YES	YES	YES
Year Fe	YES	YES	YES	YES	YES	YES
Cohort FE	YES	YES	YES	YES	YES	YES
Tenure FE	YES	YES	YES	YES	YES	YES
CohortXYear FE	YES	YES	YES	YES	YES	YES
TenureXCohort FE	YES	YES	YES	YES	YES	YES
TenureXYear FE	YES	YES	YES	YES	YES	YES
TenureXCohortXYear FE	YES	YES	YES	YES	YES	YES

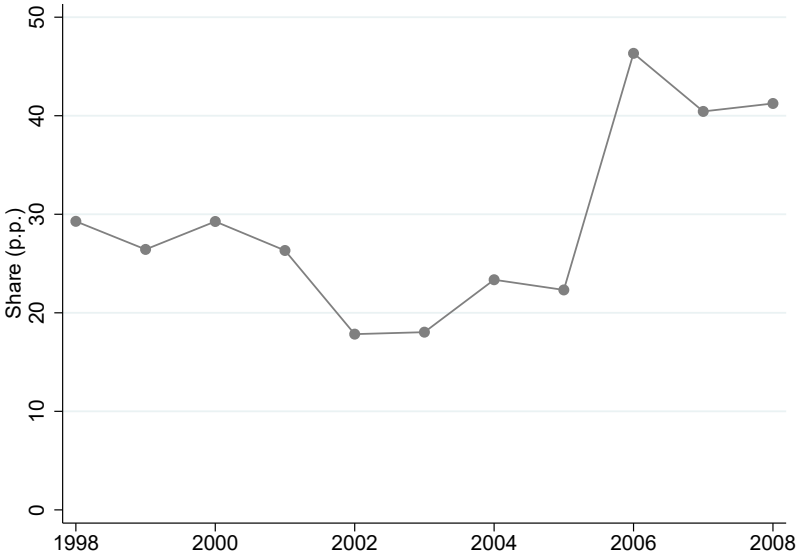
Note: able reports estimates of the effect of a change in Ofsted grade for school managers other than the principal (deputy and assistant principal) and teachers based on Equation 1. Panel covers years ranging from three years before to three years after the inspection (inspections taking place between 2006 and 2008). management and Leadership, Teaching Quality and Pupil Achievement range from 1 to 4 and are subscore for the Ofsted for specific aspects of the performance of the school. The table includes only inspections for which a change in each subscore could be computed and Column (1) reports estimates for our baseline model including only observations with available subscore changes (about half of the inspections of 2006 are lost due to absent data on subscores for the precious inspection). Tenure is computed in the year before the inspection. Standard errors are clustered at the school level at the time of inspection. Level of Significance: *** 0.01, ** 0.05, * 0.1.

Figures

Figure A1: Inspection Rating over time.



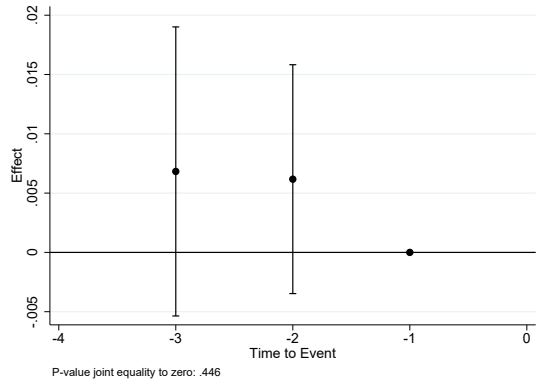
(a) Rating Change: 2003-2005 vs 2006-2008



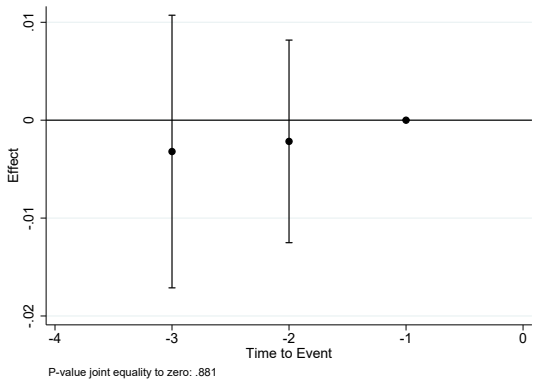
(b) Share of Schools Experiencing Downgrade

Note: Panel (a) reports inspection rating change for school which received and inspection in the period 2003-2005 and in the period 2006-2008. Panel (b) reports the share of schools which experienced a decline in their inspection rating by year of inspection..

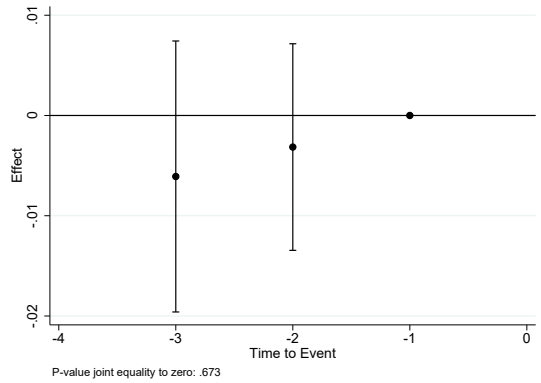
Figure A2: Differences in trends in principals' career outcome before Grade Change



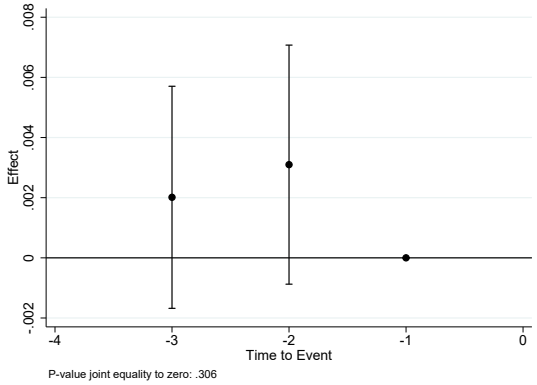
(a) (Log) Salary



(b) Principal



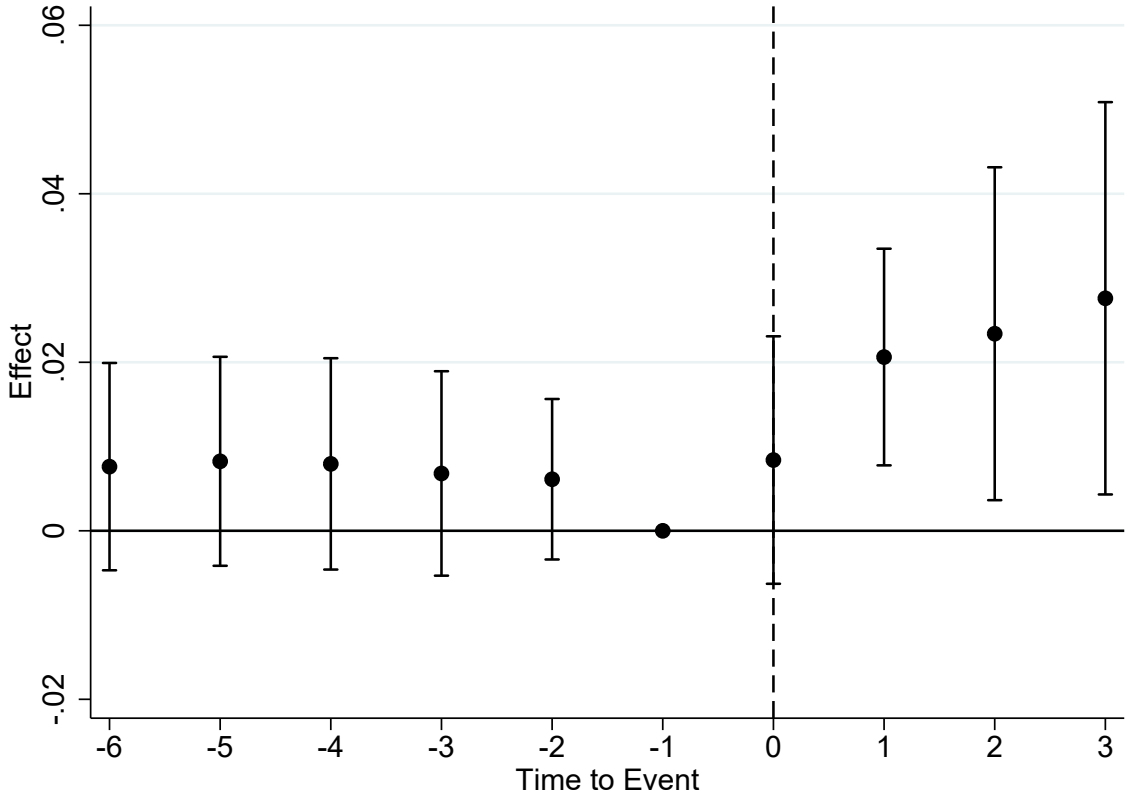
(c) Leave School



(d) Exit

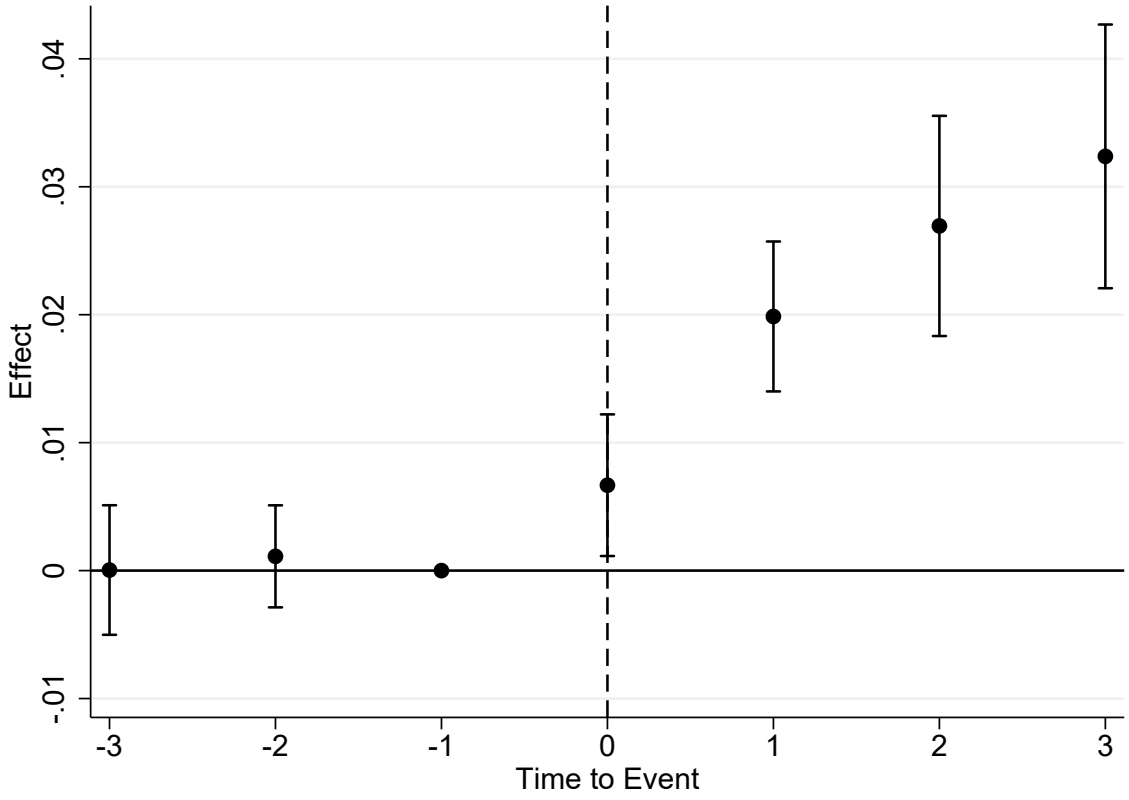
Notes: The sample is restricted to the period before the experiment. Panel (a) reports the effect on log salary; Panel (b) reports the effects on the probability of becoming a principal; Panel (c) reports the effects on the probability of changing school; and Panel (d) reports the effect on the probability of exiting the school sector. The baseline period is the year before the inspection $\tau = -1$. P-value for joint equality of coefficients to zero are reported at the bottom of each graph. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

Figure A3: Effect on salaries: Longer Pre-trends



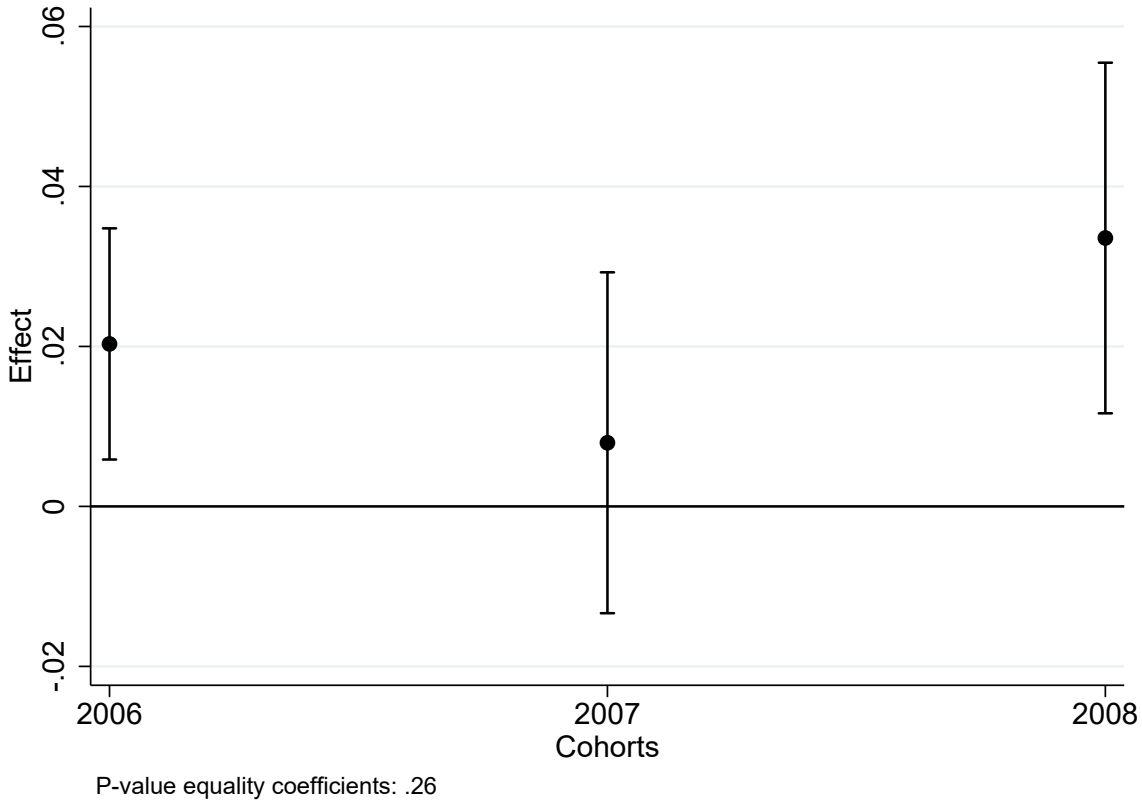
Notes: Figure reports the effect of Change in Ofsted grade on log salaries based on 2 as well as 95% confidence intervals for inspections between 2006 and 2008 with pre period extended to five years. Regression includes an age polynomial and fixed effects at individual-cohort level and at the year-cohort-tenure level. The baseline period is the year before the inspection $\tau = -1$. Standard errors are clustered at the school level at the time of inspection.

Figure A4: Effect on salaries: All years



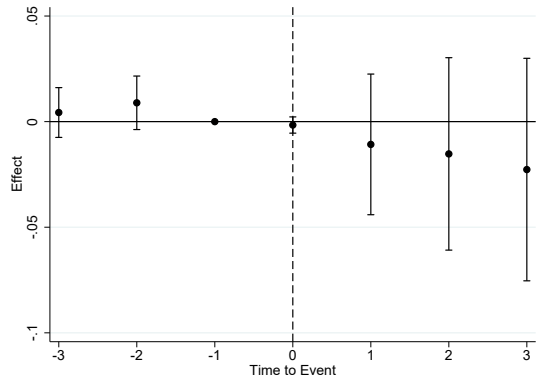
Notes: Figure reports the effect of Change in Ofsted grade on log salaries based on 2 as well as 95% confidence intervals for all years for years between 1998 and 2008. Regression includes an age polynomial and fixed effects at individual-cohort level and at the year-cohort-tenure level. The baseline period is the year before the inspection $\tau = -1$. In all cases, the regression includes years from three years prior to the inspection up to three years after the inspection. Standard errors are clustered at the school level at the time of inspection.

Figure A5: Effect of Change in Ofsted Grade on principals' Career by Year of Inspection

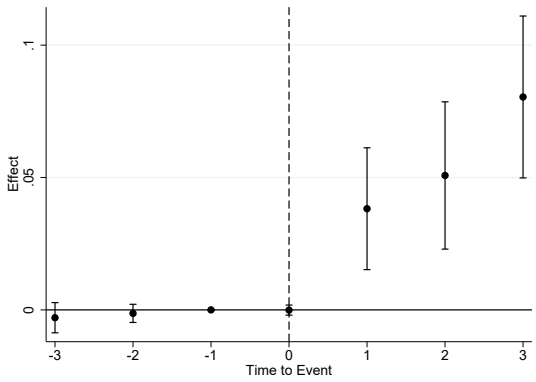


Notes: *Note*: Figure reports estimates of the effect of a change in Ofsted grade for principals' salary (Panel a) and probability of exit from the school sector (Panel b) based on Equation 1 by year of inspection, between 2006 and 2008. Panel covers years ranging from three years before to three years after the inspection. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Tenure is computed in the year before the inspection. Standard errors are clustered at the school level at the time of inspection. P-value for the equality of the coefficients reported at the bottom of the graph. Level of Significance: *** 0.01, ** 0.05, * 0.1.

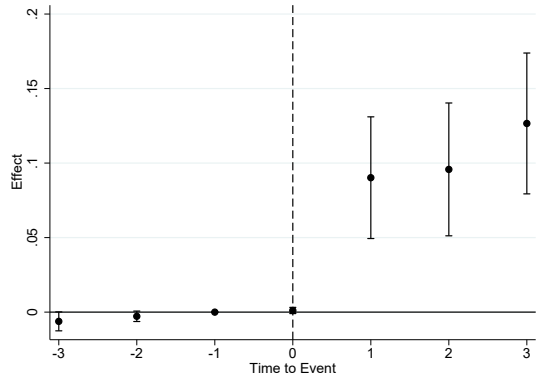
Figure A6: Effect of Change in Ofsted Grade on Exit: Improvement and Decline



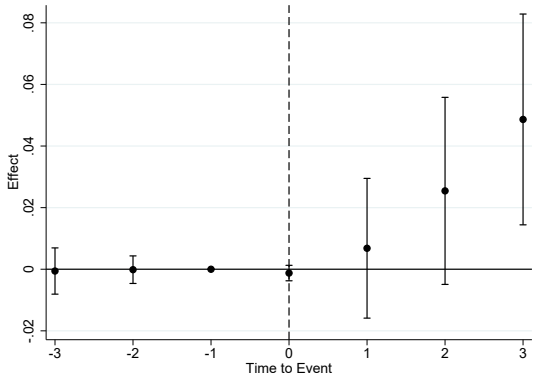
(a) Improvement



(b) Worsening



(c) Worsening - Fail

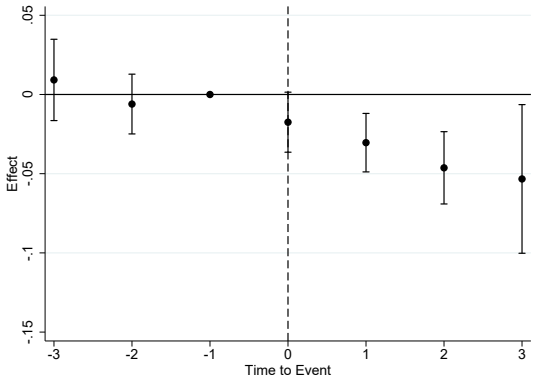


(d) Worsening - No Fail

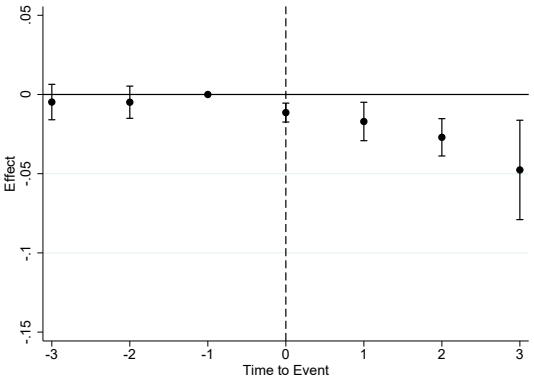
Notes: Figure reports the effect of Change in Ofsted grade on various outcomes based on Equation 2 as well as 95% confidence intervals. Panel (a) reports the effect of a grade improvement; Panel (b) reports the effect of a grade worsening; Panel (c) reports the effects of a grade worsening which does not lead to a "fail" grade for the school; and Panel (d), finally, reports the effect of a grade worsening leading to a fail grade. The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

Figure A7: Effect of a decline in Ofsted Grade on principals' (log) Salary - by initial inspection rating

Panel A: Decline - No Fail

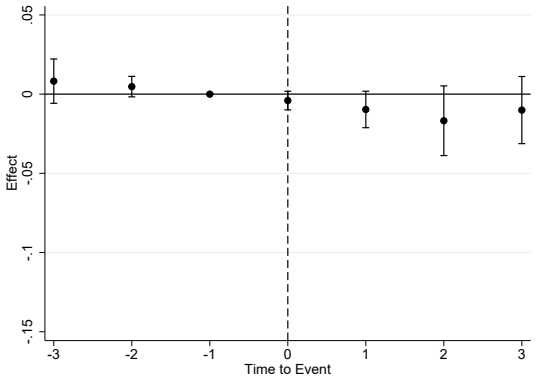


(a) Grade: "Outstanding"

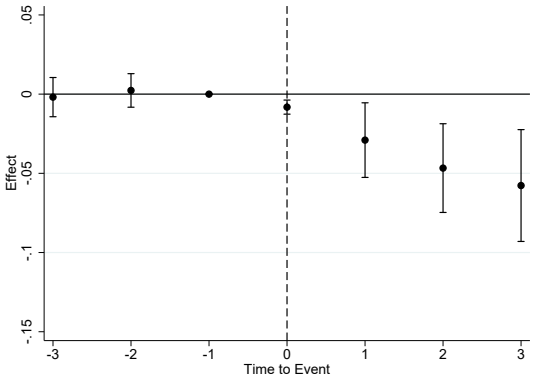


(b) Grade: "Good"

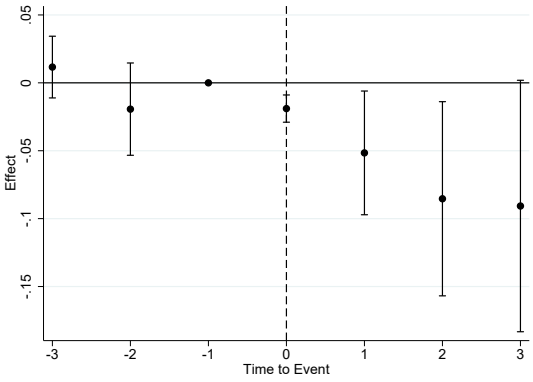
Panel B: Decline - Fail



(c) Grade: "Outstanding"



(d) Grade: "Good"

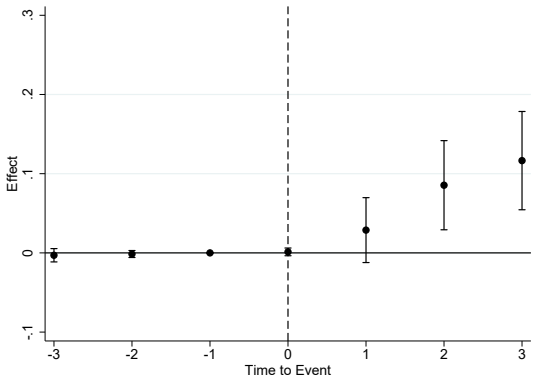


(e) Grade: "Satisfactory"

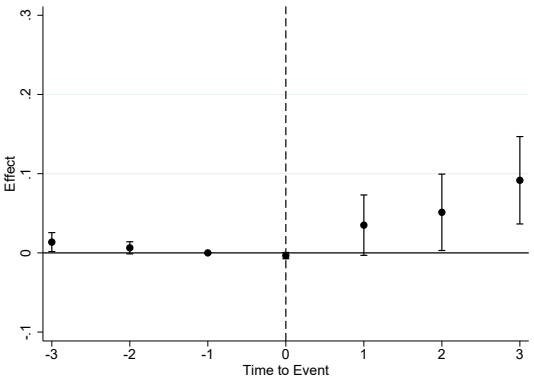
Notes: Figure reports the effect of a rating decline in Ofsted grade on principals' salary. Panel A reports the effect of a grade decline which does not lead to a "Fail" grade, while Panel B reports the impact if the grade deterioration leads to a "Fail" for the school. Panel A does not consider the "Satisfactory" starting grade, since a decline in grade would necessarily lead to a "Fail". The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

Figure A8: Effect of a decline in Ofsted Grade on principals' Exit Probability - by starting grade

Panel A: Decline - No Fail

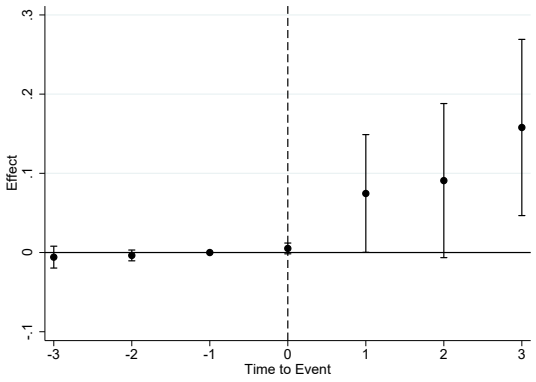


(a) Grade: "Outstanding"

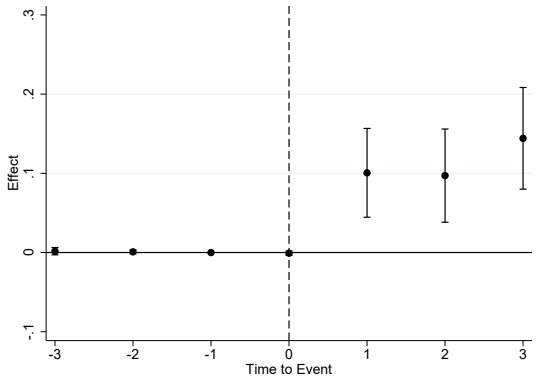


(b) Grade: "Good"

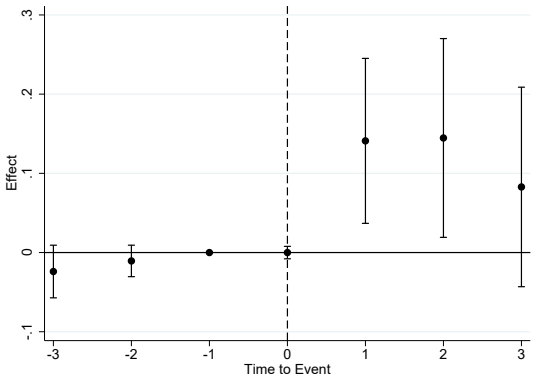
Panel B: Decline - Fail



(c) Grade: "Outstanding"



(d) Grade: "Good"

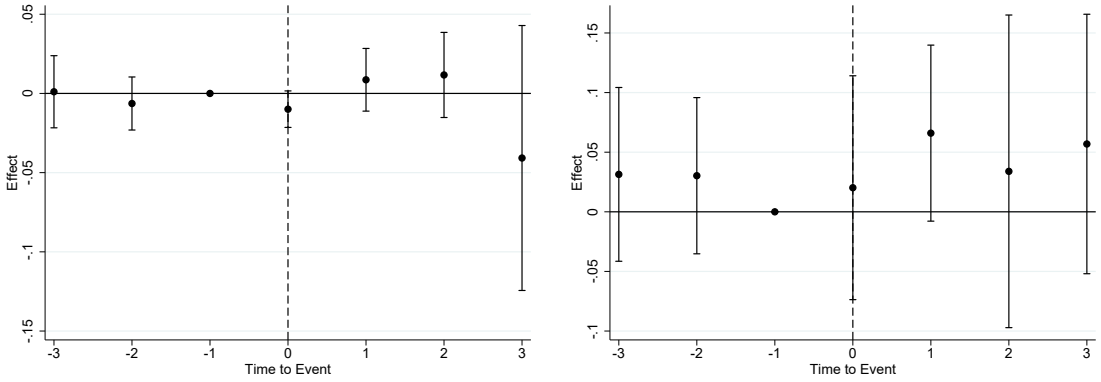


(e) Grade: "Satisfactory"

Notes: Figure reports the effect of a rating decline in Ofsted grade on principals' probability of exit from the English school system. Panel A reports the effect of a grade decline which does not lead to a "Fail" grade, while Panel B reports the impact if the grade deterioration leads to a "Fail" for the school. Panel A does not consider the "Satisfactory" starting grade, since a decline in score would necessarily lead to a "Fail". The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

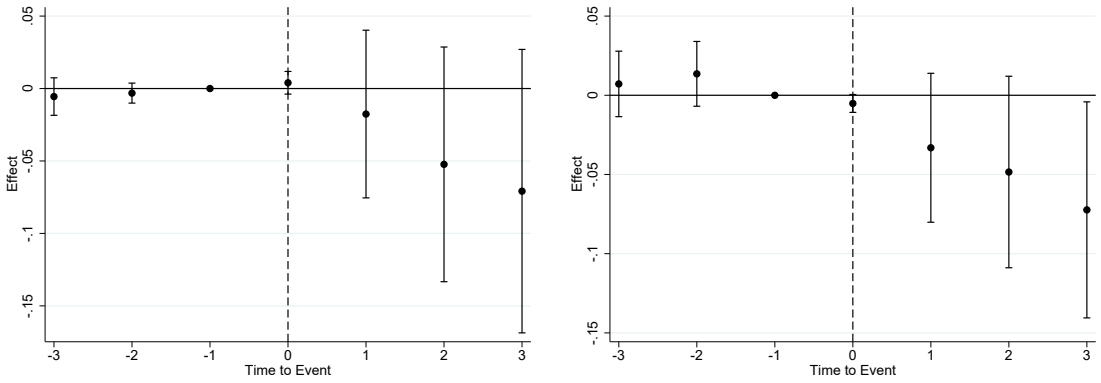
Figure A9: Effect of a rating improvement on principals' Salary and Exit Probability - by starting Ofsted grade

Panel A: (log) Salary



(a) Grade: "Good" (b) Grade: "Satisfactory"

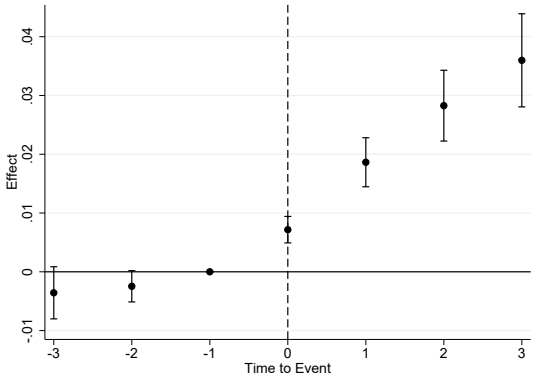
Panel B: Exit Probability



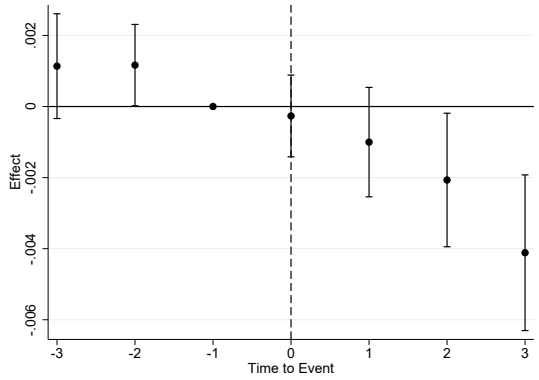
(c) Grade: "Good" (d) Grade: "Satisfactory"

Notes: Figure reports the effect of a rating improvement in Ofsted grade on principals' salary (Panel A) and the probability of exit from the English school system (Panel B). The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

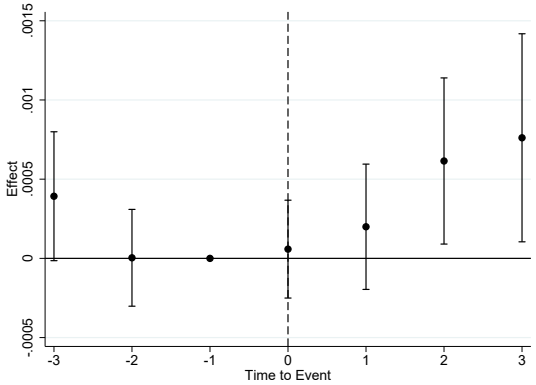
Figure A10: Effect of Change in Ofsted Grade on School Outcomes



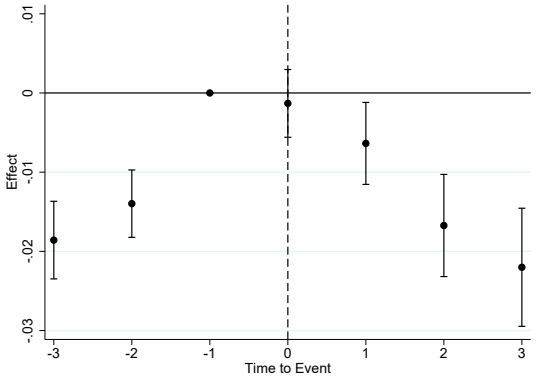
(a) (Log) Students



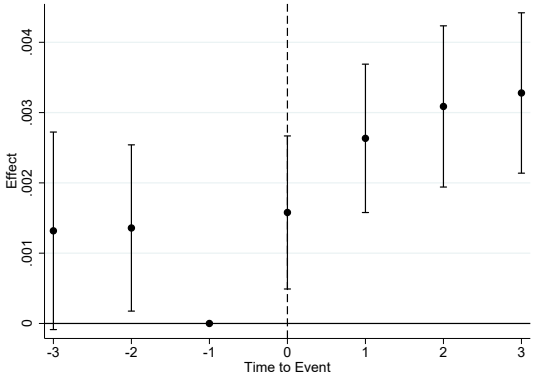
(b) % FSM



(c) % SEN students



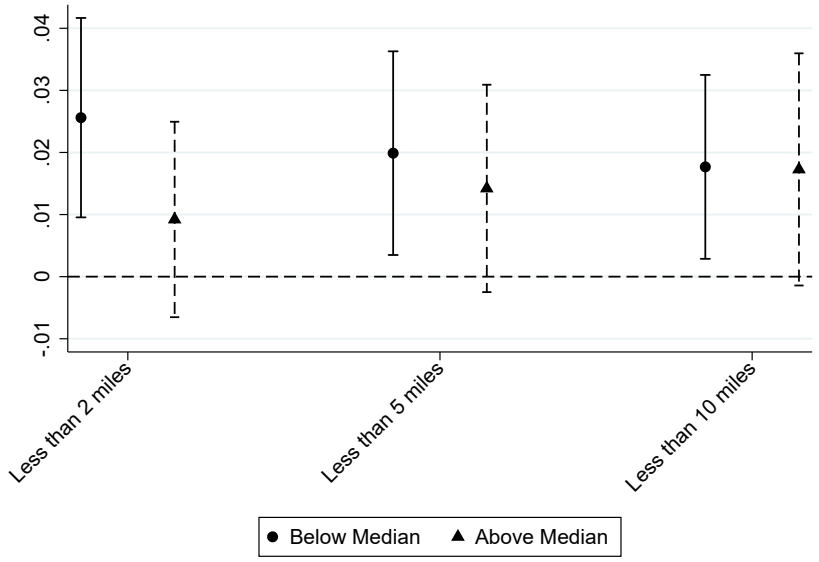
(d) % Students with 5+ A*-C in GCSE



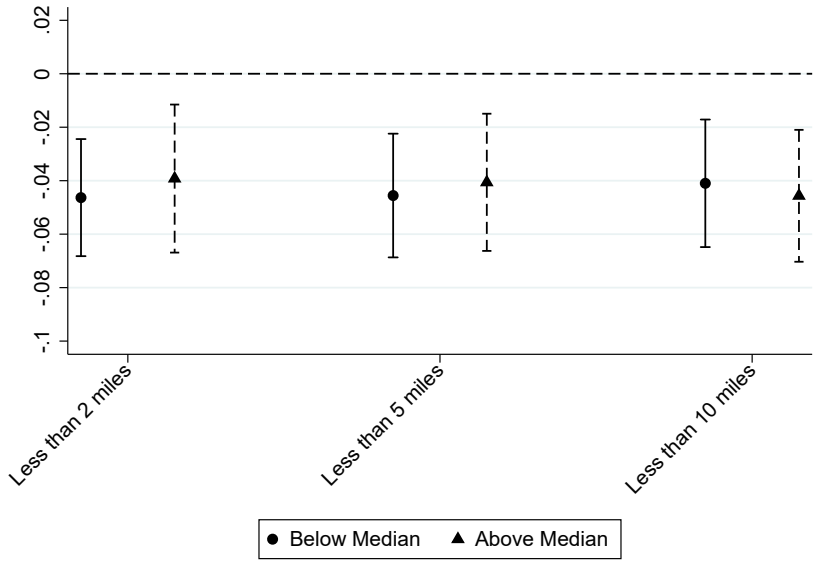
(e) % No passes (GCSE)

Notes: Figure reports the effect of Change in Ofsted grade on various school level outcomes outcomes based on Equation 2 (at school level) as well as 95% confidence intervals. Panel (a) reports the effect of a grade change on the number of pupils; Panel (b) reports the effect on the share of students eligible to free school meals (FSM); Panel (c) reports the effect on the share of students achieving 5 or more A*-C in the GCSE exam (this is considered a measure of high performance of students in the final secondary school exam, the GCSE; Panel (d) reports the effect on the share of students achieving no passes; and Panel (e), finally, reports the effect on share of students classified with Special Education Needs (SEN) with statement. The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level.

Figure A11: Number of neighboring schools and Impact of Ofsted Rating



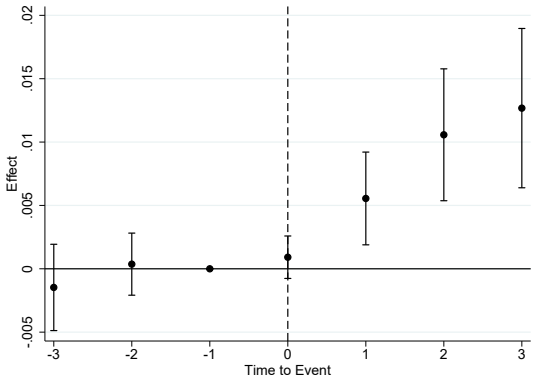
(a) Log Salary



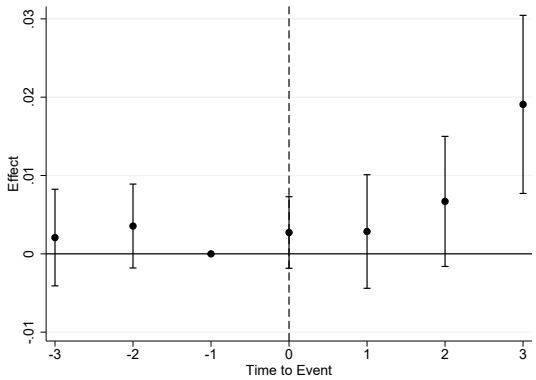
(b) Exit

Note: The figure reports coefficients and relative 95% confidence interval for the impact of a unitary change in Ofsted rating depending on the number of schools near the inspected school. Each estimate comes from a separate regression including a polynomial in age, principal fixed effects, and year by cohort by tenure at the time of the Ofsted inspection fixed effects. Panel (a) reports the impact on log salaries while panel (b) reports the impact the probability of exit. The median number of schools depends on the distance and it is: one for two miles; six for five miles; and seventeen for ten miles. Standard errors are clustered at the school at the time of the inspection level.

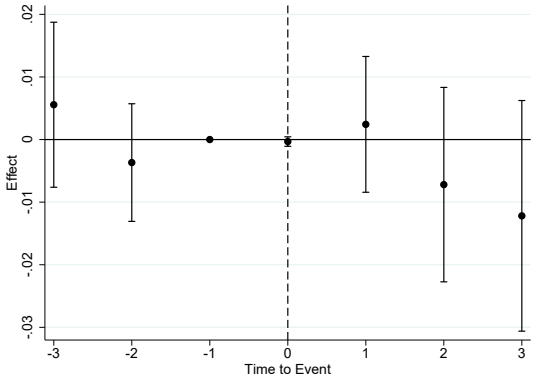
Figure A12: Effect of Change in Ofsted Grade on Senior Leaders' Salary and Career - Excluding the 2007 Cohort



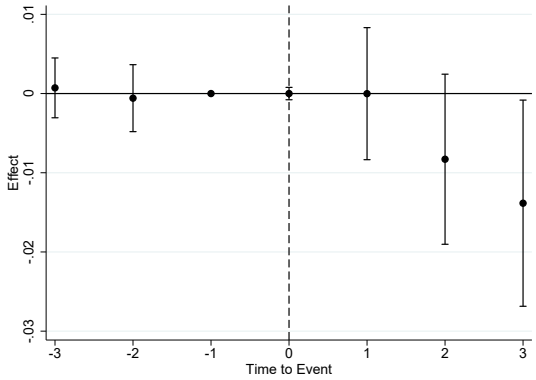
(a) (Log) Salary



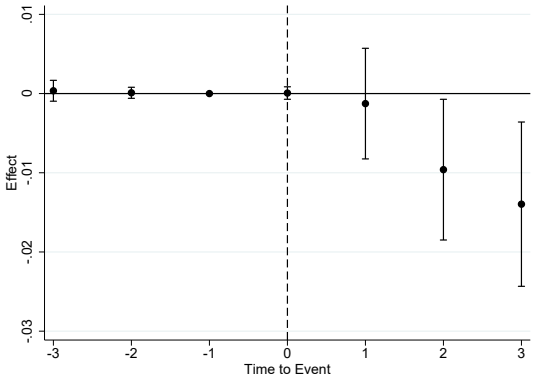
(b) Principal



(c) Leave School



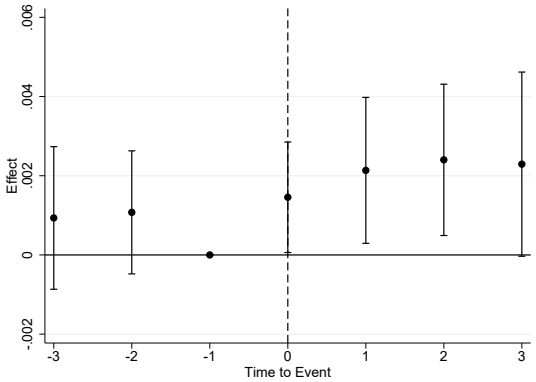
(d) Exit



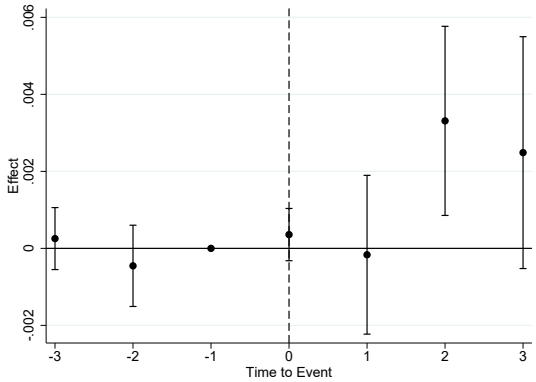
(e) Pension

Notes: Figure reports the effect of Change in Ofsted grade on Senior Leaders' outcomes based on Equation 2 as well as 95% confidence intervals. We exclude schools inspected in 2007. Panel (a) reports the effect on log salary; Panel (b) reports the effects on the probability of being a senior leader; Panel (c) reports the effects on the probability of becoming a principal; Panel (d) reports the effects on the probability of changing school; and Panel (e) reports the effect on the probability of exiting the school sector. The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level at the time of inspection.

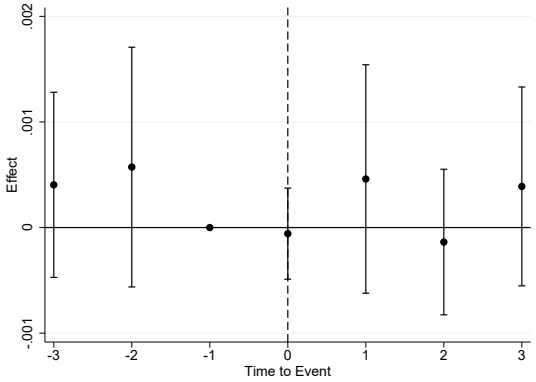
Figure A13: Effect of Change in Ofsted Grade on Teachers' Salary and Career - Excluding the 2007 Cohort



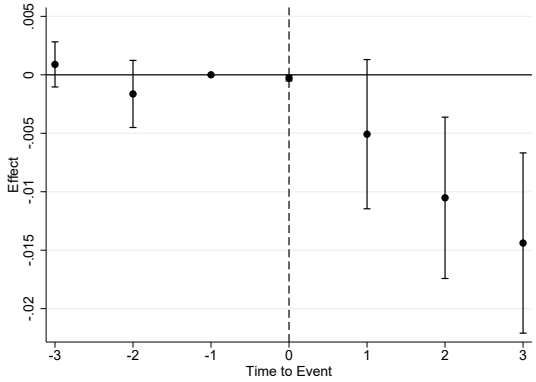
(a) (Log) Salary



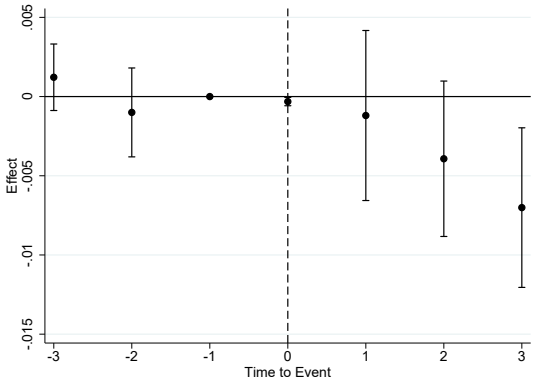
(b) Senior Leader



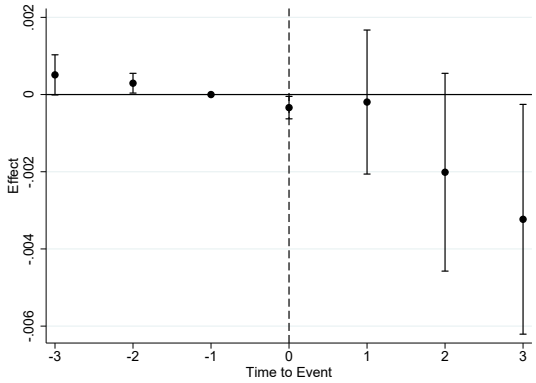
(c) Principal



(d) Leave School



(e) Exit



(f) Pension

Notes: Figure reports the effect of Change in Ofsted grade on teachers' outcomes based on Equation 2 as well as 95% confidence interval. We exclude schools inspected in 2007. Panel (a) reports the effect on log salary; Panel (b) reports the effect on salary in levels; Panel (c) reports the effects on the probability of being a senior leader; Panel (d) reports the effects on the probability of becoming a principal; Panel (e) reports the effects on the probability of changing school; and Panel (f) reports the effect on the probability of exiting the school sector. The baseline period is the year before the inspection $\tau = -1$. Regression includes an age polynomial and fixed effects at the following levels: individual by cohort, tenure at the time of the inspection, cohort, year, cohort by year, cohort by tenure at the time of inspection, tenure at the time of inspection by year, and tenure at the time of inspection by cohort and year. Standard errors are clustered at the school level, at the time of inspection.