Are Public Sector CEOs Different? Leadership Wages and Performance in Schools*

Timothy Besley
London School of Economics

Stephen Machin
University College London and London School of Economics.

February 28, 2008

Abstract

This paper uses a unique data set to investigate the link between the pay and performance of school principals. It is frequently argued that public sector CEOs are paid like bureaucrats with little reward for good performance. However, this ignores the possibility of implicit labor market incentives. We show that pay is indeed linked to publicly observable performance measures. Moreover poorly performing principals face a higher chance of being replaced. These findings for the public sector parallel what has been found by looking at the pay and performance of private sector CEOs.

*We are grateful to Joan Wilson and Ifty Hussain for excellent research assistance. We would like to thank the (then) Department for Education and Skills for part funding this project and for providing access to the administrative data we utilise and the Economic and Social Research Council for support under the Public Services Program. Thanks also go to Mono Chatterji, Rick Hanushek, Andrew Oswald, Paul Peterson, Joan Wilson and participants in a number of presentations for comments on earlier drafts of the paper.

1 Introduction

Finding ways of raising educational standards remains one of the most pressing policy agendas around the world. A series of policy initiatives are under way that change the framework relating to competition and incentives that affect the delivery of education. The US Charter school movement and the numerous education initiatives in England under the current Labour government are good examples. This high public policy profile has spawned a large body of research which evaluates the impact of such reforms. Some of this work is concerned with the way in which the development of education systems and educational policy in the last twenty years or so, including aspects of increased choice and accountability, has impacted upon the efficiency and equity of the education sector (see, *inter alia*, Hoxby, 2003, Heckman and Krueger, 2004, and Machin and Vignoles, 2005).

One frequently cited factor for improving school performance is encouraging stronger leadership from school principals (head teachers). There are many references to this in the practitioner literature on educational leadership. For example, Harris (2006) suggests in a U.S. context that "A strong, passionate principal can change a school from a good school to an excellent school" (page 28). This view has also been expressed by policy thinktanks such as the UK Policy Exchange who describe the process as "Holding out for a Hero" (UK Policy Exchange (2007, page 9) and argue that "superheads" are needed to lead under-performing schools out of failure.

If this view is correct, we should expect schools to seek ways of attracting and retaining excellent heads. Given that heads are likely to be motivated agents in the sense of Besley and Ghatak (2005) caring about mission as well as money, we would expect a combination of pecuniary and non-pecuniary incentives to be used to do this. This latter point often means giving more authority to head teachers to manage schools and shape their direction free from interference by education ministries. But equally some kind of financial rewards may be needed too. Although there is limited flexibility in public sector labour markets which lack access to stock options and explicit bonus pay, we would expect whatever instruments are available to be exploited.

The past twenty years have witnessed a big increase in the study of corporate governance and the role of Chief Executive Officers (CEOs) in the management of firms (see Jensen and Murphy, 1990, and the review of Murphy, 1999). It is now well understood that even when firms are subject to competitive pressures, the structure of governance matters to firm per-

formance. There can be little doubt that leadership in the private sector matters and this shows up in the way in which CEOs are rewarded.

A striking feature of the existing literature is the development of a wide variety of data sets that are used to look at the way CEOs are remunerated and determinants of CEO turnover. This has been connected to the performance of the organizations for which they work, with an intention of developing tests of simple incentive models that have predictions on the way in which executives are rewarded. The evidence base resulting from this work now illuminates academic and policy literatures on corporate governance - see, for example, Murphy (1999), Hall and Leibman (1998) and Bertrand and Schoar (2003).

The existing empirical literature on CEO pay and performance is almost exclusively focussed on private firms. Indeed, there is an implicit assumption that private sector CEOs should be contrasted with public sector "bureaucrats" whose pay and performance are not linked. There are *a priori* grounds to be suspicious about this claim even though explicit bonus pay is rarely used. Public sector CEOs are recruited in open competition where recruitment and retention is a vital issue. It is well-known from the classic career concerns model of Holmstrom (1999) that implicit (market based) incentives can be important in relating pay and performance. Issues of leadership are equally important in the public sector, not least because market pressures are usually much weaker. Yet there is almost no literature amongst economists on the study of public sector leaders.

Looking at the role and impact of key players in the education sector ties in well with the recent expansion of work on the economics of education (see Machin and Vignoles, 2005, or Machin, 2008). This paper takes advantage of a unique data set to look at these issues. We use a matched data set comprising of all high schools and their principals in England (secondary schools and their head teachers in local language) between 1994 and 2002. The data set includes the salary paid to each head teacher during this period. We take advantage of the fact that since 1992 the government published league tables to publicize the performance of schools on a commonly measurable basis, essentially scores in nationally administered exams taken at age 16. This provides a clean measure of school performance that is deemed to be highly relevant to most parents and school governors. We ask of the data whether there is a robust link between pay and school performance. We also investigate determinants of turnover among heads.

The remainder of the paper is organized as follows. In the next section,

we discuss our theoretical issues as we set out a simple model of learning in the labor market for head teachers. This motivates our view of the pay-performance gradient and turnover in the education sector. Section 3 of the paper describes the data we use, drawing on administrative data on head teachers in secondary schools in England. Section 4 lays out the empirical model and presents results that try to evaluate the key predictions and insights of our theoretical model. Section 5 concludes.

2 Theoretical Issues

In this section we discuss the underlying theoretical reasons to expect a link between school performance and teacher pay.

2.1 Basic Model

The basic model is based on the career concerns model of Holmstrom (1999), which was used to look at the pay-performance relationship by Gibbons and Murphy (1992). However, for simplicity, we abstract from any effort decision by head teachers and focus purely on screening based on past performance. Implicitly, therefore, we are appealing to the fact that head teachers are intrinsically motivated. Head teaching careers last for two periods. We assume that all head teachers are productive in period one of their careers. However, there is a probability γ that a head will not be productive in period two. While crude, this allows us to capture in reduced form the possibility of a career/effort profile.[1] Over time the market will reward head teachers based on their revealed performance.

Suppose that head-teachers are matched with a school s in each time period. Each head teacher is of ability α_i which is drawn from a normal distribution with mean μ and variance σ_α^2. All head teachers are *ex ante* identical.

The production function for school quality with head teacher i in school s at date t is:

$$q_{ist} = (\alpha_i + \varepsilon_{ist})\,\delta + \theta_s$$

[1] It would be possible to offer a micro-foundation for this in terms of effort, in which a fraction $(1-\gamma)$ of heads are intrinsically motivated and always put in effort while a fraction γ are opportunists who put in no effort in the second phase of their careers.

where $\delta = 1$ if the school head is productive and ε_{ist} is an idiosyncratic shock to head teacher productivity which is normally distributed with mean zero and variance σ_ε^2. The variable θ_s captures variation in school productivity due to factors such as neighborhood, parental involvement etc. We treat this as a permanent (non time-varying) effect and assume that the distribution of school productivity across schools is drawn from a normal distribution with mean $\bar{\theta}$ and variance σ_θ^2. The variables $(\alpha_i, \varepsilon_{ist}, \theta_s)$ are assumed to be independently distributed over time and across schools/teachers.

Those active in the market for head teachers observe quality at the school at which the head is working, q_{ist}, and the background characteristics of the school in question, θ_s. Observing the latter is important since otherwise the head teacher's reputation would be built in part on drivers of school quality rather than their own contribution to school performance. Thus, the market judgement on the quality of the head will offset any performance of heads due to underlying school quality. Using standard results for normal distributions, it is straightforward to see that:

$$E\left\{\alpha : q_{ist}, \theta_s\right\} = \mu + \beta\left(q_{ist} - \theta_s\right) \tag{1}$$

where

$$\beta = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$$

is the regression coefficient relating ability on the head teacher specific input to school quality.

In the basic model, suppose that the head teacher is paid their marginal product based on their expected ability inferred from their performance. This implies that:

$$w_{ist} = \gamma\left(\mu + \beta\left(q_{ist} - \theta_s\right)\right).$$

The parameter γ reflects discounting due to the fact that some head teachers will become productive in period two of their careers.

This model has the standard empirical implication that performance of the school and wages should be related. To see this, suppose that we estimate the following regression linking wages and school performance:

$$w_{ist} = a + bq_{ist} + \eta_{ist}. \tag{2}$$

assuming that η_{ist} is uncorrelated with q_{ist}. Then

$$\begin{aligned}\hat{b} &= \frac{cov\left(w_{ist}, q_{ist}\right)}{var\left(q_{ist}\right)} \\ &= \gamma\beta\left(\frac{\sigma_\alpha^2+\sigma_\varepsilon^2}{\sigma_\alpha^2+\sigma_\varepsilon^2+\sigma_\theta^2}\right) \\ &= \gamma\frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\varepsilon^2+\sigma_\theta^2}.\end{aligned}$$

This says that wages and school quality should be positively correlated in so far as head teacher ability explains school quality. Indeed, $\frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\varepsilon^2+\sigma_\theta^2}$ is the regression coefficient of teacher ability on school quality. The dependence of wages on school quality is smaller the more likely it is that head teachers shirk later in their careers as represented by γ.

2.2 Incorporating Turnover

In the basic model all head teachers continue to be employed by schools regardless of their performance earlier in their careers. However, their wages move up and down over time to reflect their performance. As we shall see below, head teachers in our data turnover when school performance is poor. We now incorporate this into the theoretical analysis.

Suppose now that all head teachers have a reservation wage u which can be thought of as the wage from working as a rank and file teacher. We assume that there is a fixed stock of N schools and a supply of $M > N$ untried head teachers who can be hired at u and who have expected ability μ. A school can then compare employing an experienced head with ability level $E\left\{\alpha : q_{ist}, \theta_s\right\}$ with going to the pool of new head teachers. Since head teachers cannot be paid less than u, this implies that only head teachers with expected ability exceeding μ will remain as head teachers. Using (1), a teacher will therefore stay employed if:

$$\gamma\left(\mu+\beta\left(q_{ist}-\theta_s\right)\right) > \mu$$

or

$$q_{ist} > \hat{q}_s = \theta_s + \frac{\left(1-\gamma\right)\mu}{\beta\gamma}, \tag{3}$$

i.e. if their school has performed above a threshold level of quality. A higher value of β lowers the performance threshold as there is a stronger signal

about teacher ability in good school performance. The threshold is also decreasing in γ – the critical performance threshold is higher when it is more likely that period two head teachers will be productive. In a multi-period model we might expect γ to fall over time, in which case teachers would be held to higher standard and turnover (controlling for school performance) would increase as head teachers age. Observe from (3) that the rule for retaining head teachers is school specific, with the performance threshold being higher in better quality schools.

Thus the model predicts that the probability of a head teacher who produces quality q_{ist} in a school of type θ_s being fired is:

$$1 - \Phi\left(\frac{q_{ist} - \theta_s - (1-\gamma)\,\mu}{\beta\left(\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\theta^2\right)\gamma}\right) \tag{4}$$

where $\Phi(\cdot)$ is the distribution function of the standard normal.

We now determine the equilibrium wage of a head teacher with reputation $E\{\alpha : q_{ist}, \theta_s\}$ in a competitive labour market. Suppose that all schools care about expected quality net of head teacher wages. This implies that in a competitive equilibrium, a school must be indifferent between hiring a head teacher of any quality level. Given that a school also has the option of hiring an inexperienced head, this implies that wages for experienced heads with $E\{\alpha : q_{ist}, \theta_s\}$ solves:

$$\theta_s + \gamma E\{\alpha : q_{ist}, \theta_s\} = \theta_s + \mu - u. \tag{5}$$

On this basis, experienced heads teachers with reputations exceeding μ will earn a premium. Solving (5) implies that the equilibrium wage function will be:

$$w_{ist} = \begin{cases} u + \gamma E(\alpha : q_{ist}, \theta_s) - \mu & \text{if } \gamma E(\alpha : q_{ist}, \theta_s) \geq \mu \\ u & \text{otherwise.} \end{cases}$$

Now consider what is estimated if we run a regression of the form:

$$w_{ist} = a + b q_{ist} + \eta_{ist} \tag{6}$$

assuming once again that η_{ist} is uncorrelated with q_{ist}. In this case, the

regression estimate is:

$$\begin{aligned}
\hat{b} &= \frac{cov\left(w_{ist}, q_{ist} : q_{ist} \geq \theta_s + \frac{(1-\gamma)\mu}{\beta\gamma}\right)}{var\left(q_{ist}\right)} \Pr ob\left(q_{ist} \geq \theta_s + \frac{(1-\gamma)\mu}{\beta\gamma}\right) \qquad (7)\\
&= \delta\beta \frac{cov\left(q_{ist} - \theta_s, q_{ist} : q_{ist} \geq \theta_s + \frac{(1-\gamma)\mu}{\beta\gamma}\right)}{var\left(q_{ist}\right)} \Pr ob\left(q_{ist} \geq \theta_s + \frac{(1-\gamma)\mu}{\beta\gamma}\right)\\
&= \frac{\gamma\beta}{(\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\theta^2)}\left[var\left(\alpha_i + \varepsilon_{ist} : \alpha_i + \varepsilon_{ist} \geq \frac{(1-\gamma)\mu}{\gamma\beta}\right)\right] \Pr ob\left(q_{ist} \geq \theta_s + \frac{(1-\gamma)\mu}{\beta\gamma}\right)\\
&= \delta\frac{\gamma\beta\left(\sigma_\alpha^2 + \sigma_\varepsilon^2\right)}{(\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\theta^2)}\lambda\left(\frac{(1-\gamma)\mu}{\beta\gamma\left(\sigma_\alpha^2 + \sigma_\varepsilon^2\right)}\right)\\
&= \frac{\gamma\sigma_\alpha^2}{(\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\theta^2)}\lambda\left(\frac{(1-\gamma)\mu}{\gamma\sigma_\alpha^2}\right),
\end{aligned}$$

where $\lambda(z) = \left(1 - \frac{\phi(z)}{1-\Phi(z)}\left[\frac{\phi(z)}{1-\Phi(z)} - z\right]\right)[1 - \Phi(z)] \in [0,1]$ using well-known results for the standard normal distribution.

Comparing the estimates of the wage/school quality gradient with and without turnover, it is clear that selection will lead to a smaller correlation between performance and ability, since now poor performance leads to a head teacher being fired rather than working for a lower wage. This makes intuitive sense as we have created a wage floor for poorly performing heads. In a model where firing is made more costly, this would tend to reduce the sensitivity of wages to performance as it would reduce the scope for schools to reward performance. Thus protection of head teachers would naturally lead to the standard view of bureaucratic wage determination, in which the link between wages and performance is low or non-existent.

2.3 Match Specific Utility

The basic model above abstracts from the possibility that there are any other factors beyond head teacher ability which affect exam results and therefore the allocation of head teachers to schools. However, in practice, other factors are likely to be important. Schools have many objectives other than success in exams and head teachers are likely prefer to work in schools where they are sympathetic to the school ethos.[2] Thus we expect there to be an underlying matching process which can lead to either a positive or negative wage

[2] See Besley and Ghatak (2005) for a model along these lines.

premium depending on the extent of the match specific utility. In general, we can denote this by u_{ist} which could, in general, be correlated with θ_s and α_i. It will enter as an additional error term in the wage equation.

To see how such considerations affect the wage-performance relationship, we revert to the basic model without turnover. Thus wages are:

$$w_{ist} = \mu + \beta (q_{ist} - \theta_s) + u_{ist}.$$

In this case the regression coefficient relating wages to school quality is:

$$\hat{b} = \gamma \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\theta^2} + \beta \frac{cov(\theta_s + \alpha_i, u_{ist})}{\sigma_\alpha^2 + \sigma_\varepsilon^2 + \sigma_\theta^2}.$$

This second term represents any correlation between wages and performance due to how match specific utility is correlated with school and head teacher quality. Failing to account for this could lead us to either under or over-estimate the "true" wage-school performance gradient. The direction of this bias, however, is not clear *a priori*. If good schools are more demanding due to dealing with pushy parents and governors, we might imagine a positive covariance. However, if being a head-teacher in a good school is more prestigious and hence commands a lower wage premium, the bias in the wage-school performance relationship could be negative.

Suppose, however, that we write the match specific utility terms as follows:

$$u_{ist} = \rho_s + \rho_i + \rho_{ist},$$

i.e. where we have decomposed these idiosyncratic wage terms into a set of school specific factors, ρ_s, head-teacher specific factors, ρ_i, and an idiosyncratic (white noise) component ρ_{ist}. Then, incorporating school and head teacher fixed effects will help to clean out some of the sources of bias that we might observe in estimating the wage-performance gradient. This turns out to be important in our empirical analysis and we return to it in discussing the results below.

3 Data Description

3.1 Data Sources

The data we use is a very rich administrative data source on teachers in English schools that contribute to the Teachers Pension Scheme, known

as the Database of Teacher Records (DTR). We look at head teachers in English state secondary schools (attended by children aged 11 through 16) over the years 1994-2002.[3] There are approximately 4,000 observations on head teachers per DTR year. We concentrate on head teachers currently in service.

We need to match DTR data on head teachers to school level data and to do so we established a school identifier based on DTR records of the school establishment number and Local Education Authority code for each year. This identifier was used to match in annual school performance tables statistics, such that our data set encompasses an array of key indicators of school characteristics. These include measures of pupil performance at the end of secondary school, total pupil numbers and school types (e.g. whether community, foundation or voluntary).

In selecting only head teachers from the DTR data set we found some cases where schools had more than one head teacher observation for the year. This required us to choose the head teacher we perceived was most likely to have been in leadership in that year for that school, such that we had only one head teacher observation per year. Manual selection of the head teacher observation took place for 2,430 schools and was based on a range of indicators: salary information (highest salary), tenure (longest tenure), and last recorded service date.

The data set includes generated variables to indicate head teacher turnover and employment activity, where the former includes turnover both within and between the years and the latter establishes the prior employment status of those head teachers that are not in our sample from the beginning (1994). Through this we are able to establish promotional activity - that is whether a current head teacher was promoted from within the same school or another school and from which teaching level they were promoted – and which head teachers are entirely new to the profession.

Because we do not end up with matched data on head teachers and school performance in all years (for a number of reasons like missing performance data, a small number of schools merging or closing, and missing head teacher data) we present results throughout for a balanced panel of 1790

[3]The way that the head teacher coding variable is collected was changed after 2002, making it much harder to identify head teachers. This is the reason our sample stops in 2002. This said, we have attempted to extend the sample to 2005 and, when we did so, obtained similar results, but with a drop off in the number of matched head teachers and schools. Hence we stick with the 1994 to 2002 sample in this paper.

schools and for an unbalanced panel, where we specify the need for at least four continuous time series observations. This unbalanced panel covers 2601 schools over the nine years we study.

3.2 Descriptive Statistics

Some descriptive statistics are given in Table 1. The upper section of the Table shows descriptives for the unbalanced panel and the lower part for the balanced panel. Annual means are given for four variables: head teacher salary; the headline figure of secondary school performance usually considered, namely the proportion of students obtaining 5 or more grade A*-C examination results in the final year 11 exams (the General Certificate of Secondary Education, GCSE); the proportion of students getting no GCSE passes; and the proportion of head teacher turnovers.

Head teacher salaries are set by school governors with reference to the leadership spines determined by the School Teachers' Review Body. The salaries are linked to school size groups, but there is also scope for governing bodies to pay more where necessary to recruit and retain teachers. Thus schools (and local authorities) do have flexibility to move teachers up the pay spines and to set a head teacher's salary above the top of the pay spine. In fact, looking at our data suggests that in a given year around 20 percent of heads are above the salary spine one would expect them to be on, showing scope for incentive payments to be used for heads.

It is worth discussing the school performance measures. The GCSEs are standardized national exams taken by all pupils in England in the last year of compulsory schooling at age 16. A GCSE examination mark is deemed a 'pass' if it falls within the grade range A*-G. A grade below a G is a fail. A 'good pass' is in the A*-C grade range and, as already noted, the government headline figure measures the proportion of pupils getting five or more of these good passes. A key advantage for our empirical analysis is that these are well known, publicly available measures of school performance. The particular measures we use are dictated by their availability in the published league tables.[4]

[4]One should note that, whilst the headline figure we utilise has been published in the publically available league tables throughout the time of their existence, there have been heated debates about: i) the usefulness of league tables; and ii) what are the 'right' measures of performance. Because of this their publication format has evolved over time, the most notable innovation being the publication of value added numbers in addition to

The upper section of Table 1 shows the average head teacher annual salary to be around £58,000 in 2002, having risen from just under £40,000 in the first year that we consider, 1994. The headline school achievement figure, the proportion of children getting 5 or more A*-C GCSEs, rises from .39 in 1994 up to .48 by 2002. The measure of poor performance, the proportion of pupils with no GCSE passes, also shows measured school performance to be improving through time, as it falls from .07 to .04 between 1994 and 2002. Comparing these numbers from the unbalanced panel to the balanced panel numbers in the lower part of the Table is highly reassuring – the means are almost identical.

Moving next to consider head teacher turnover, the Table shows that around one in ten head teachers leave their job in a given year, although there are some year-to-year fluctuations in the turnover rate. Again the Table shows a reassuringly similar pattern for the balanced and unbalanced panels of teachers and schools that we study here.

4 Results

4.1 Method

In this section we use the DTR data to consider the predictions arising from the model presented in Section 2. We first consider the correlation between head teacher wages and school performance by estimating a pay-performance equation like those estimated in the sizable literature on CEO pay in the private sector (for example the empirical analysis of US chief executives by Jensen and Murphy, 1990).

The estimating equation relates the log of head teacher salary to school performance between 1994 and 2002. The starting point is simply a regression on the pooled data of the following form for head teacher i in school s in year t:

$$\ln(w_{ist}) = D_t + bq_{ist} + cX_{st} + \eta_{ist} \tag{8}$$

where w_{ist} is the head teacher's salary in school s in year t, D_t are a set of year dummies, q_{ist} is school performance, X_{st} are time-varying school charac-

the headline figure. These value added numbers were first published in 2005, which is after our estimation period. The other point of relevance is that the league tables are probably also used more heavily by parents who are more 'able' to interpret their meaning.

teristics and and η_{ist} is a white noise error term. The year dummies capture common changes in school performance due to (say) external interventions to improve quality and head teachers' pay and conditions. The main time varying characteristic that we include, following the extensive CEO literature (see the literature survey in Murphy, 1999), is the size of the school.

As we discussed in section 2.3, one concern with estimating (8) is that there are school characteristics which make teaching in a specific school more or less challenging or pleasant. These are difficult to measure and may be correlated with θ_s, which affects school quality. One way to deal with this is to include school fixed effects D_s and to estimate:

$$\ln(w_{ist}) = D_t + D_s + bq_{ist} + cX_{st} + \eta_{ist} \tag{9}$$

Finally, we allow for the possibility that there are head teacher characteristics which may be correlated with their ability to generate school quality, making them attractive to schools and affecting their pay. If these characteristics are fixed then we can deal with this by including a set of head teacher fixed effects D_i to estimate:

$$\ln(w_{ist}) = D_t + D_s + D_i + bq_{ist} + cX_{st} + \eta_{ist}. \tag{10}$$

In this case the effect of school quality on wages is identified from deviations from school and individual teacher means. The results that follow compare all three specifications that we have discussed here.

4.2 Teacher and School Fixed Effects

The nature of our data means that, in an analogous manner to the large literature on matched worker-firm data (Abowd and Kramarz, 1999), we can begin by considering the relative importance of head teacher and school fixed effects in some basic salary equations. This is a useful descriptive device for understanding the data and to offer a precursor to the more specific salary models we use to test the theoretical model presented above.

To assess the relative importance of the different fixed effects, we just carry out a basic variance decomposition type of analysis of log(salary) in a regression like (9) that, for now, excludes the q_{ist} and X_{st} variables. This is rather like the exercise undertaken by Bertrand and Schoar (2003) who focus on the relative importance of individual manager and firm fixed effects in explaining firm performance.

The results are shown in Table 2, which reports six specifications each for the balanced and unbalanced panels of teachers and schools. The first three specifications exclude the year effects D_t , whilst they are included in the final three. The two sets of three specifications include only head teacher fixed effects (columns (1) and (4)), only school fixed effects (columns (2) and (5)) and both (columns (3) and (6)). The results in the Table make it clear that the there are important head teacher and school fixed effects, as shown by the F tests of joint significance. Moreover, the contribution of head teacher fixed effects to the variance of salary is largest.[5]

From this we conclude that there are significant salary differences between the head teachers in our sample, over and above the school in which they work. We next focus on the relationship between salary and school performance measures.

4.3 Core Results

4.3.1 Pay-Performance Estimates

Table 3 reports estimates of the salary equations (8) through (10). The table is structured so as to build up from the simplest model (the pooled model, equation (8)) by adding the various fixed effects and ending up with the most detailed model incorporating both school and head teacher fixed effects (equation (10)).

Column (1) of Table 3 reports estimates of the basic pooled model in (8). This finds a positive and significant estimate on the headline school performance measure based on exam scores. This suggests, in line with the basic idea of the model, that salaries of head teachers are associated with higher observed school performance, i.e. leaders of better performing schools are paid more.

Column (2) of Table 3 shows the result of adding a full set of school fixed effects to the pooled model given in column (1). This specification studies the relationship between changes in salaries and performance within schools over time. The positive correlation between salary and performance observed in columns (1) is confirmed in this specification although the pay-for-performance sensitivity, whilst remaining strongly significant, is almost

[5]See, for example, the R-squared of the regression increasing by more on inclusion of the head teacher fixed effects.

halved by the inclusion of the school fixed effects.[6]

Column (3) of Table 2 adds school size to the column (2) specification. It shows that school size is also a strong determinant of head teacher salaries. This is not surprising given that head teacher salary scales are directly linked to the size of the school that they head up. But it is interesting that the association with school performance remains positive and significant. Over and above school size we are able to uncover a significant association between head teacher salaries and pupil performance. This result also offers an interesting parallel to the private sector CEO literature, where size (in levels and differenced models) is also a key determinant of compensation and one that is robust across many studies (Murphy, 1999).

Column (4) further refines the school performance measure, by also adding in a measure of poor performance, namely the proportion of children failing at GCSE level, i.e. getting no passes whatsoever. The coefficient on this variable is negative and significant, while the estimated coefficient on the headline exam score measure remains significant and positive (and of very similar magnitude). Thus head teachers are rewarded for good performance relative to the school mean and punished financially for poor performance.

The pay-performance gradients are thus statistically significant, showing that efforts to recruit or retain good heads are at work for head teachers. In terms of magnitudes, model (4) shows a head teacher in a school with all students getting the headline figure of 5 or more A*-C GCSEs receives a pay premium of about 3.2 percent relative to a head in a school where no students do so and all get at least some passes lower than a C grade, and a sizable 9.8 percent compared to a school where all students get no GCSE passes. Of course, this compares salaries in the best performing relative to the worst performing schools, but the size of the salary gaps is economically important.

In the final three columns of Table 3, (5), (6) and (7), we additionally include the head teacher fixed effects thereby estimating the most general model our data permits. The positive salary-performance gradient predicted by theory remains intact when head teacher fixed effects are included. Head teacher salaries and school performance do seem to move together, confirming the importance of a pay-performance gradient of the kind that has been

[6]In addition, conditioning on head teacher experience in the teaching profession (i.e. controlling for date of entry to the profession) leaves these results essentially unchanged. As the experience variable drops out of the more detailed models that include head teacher fixed effects, we report results that exclude this experience variable.

stressed in the literature on private sector CEOs. The 3.2 and 8.9 percent full range effects for the school fixed effects only models fall a little, but now become 2.2 and 7.6 percent respectively. Calculating magnitudes more conservatively, an increase in the headline 5 or more A*-C GCSEs measure from .4 to .8 (around 2 standard deviations) in the balanced panel results generates a .8 to 1.2 percent increase in salary. Increasin the no GCSE passes measure by .1 results in a salary fall of .8 to 1 percent. The effects are thus quite modest and compared to average salaries are only of the order of a few hundreds of pounds. This is not surprising in the context of the public sector labour market we study and, in fact, would probably be counterintuitive were they much larger. Nonetheless, they still show evidence of incentives at work for head teachers.

Overall, the empirical results accord well with the theoretical model which suggests that the pay-performance gradient comes from learning about which head teachers are effective in delivering school quality. The school and head teacher fixed effects should go some way towards dealing with match specific factors in pay determination. The fact that the estimated coefficient on school performance falls when school and head teacher fixed effects are included suggests that there are omitted school and head teacher characteristics that are positively correlated with one another and hence lead to an upward bias in the pay-performance relationship. In terms of the discussion of match-specific utility in section 2.3, this suggests that $cov\left(\theta_s + \alpha_i, u_{ist}\right) > 0$, which would be true, for example, if head teachers get a positive utility level from working in good schools.

Taken together, the results in Table 3 suggest a statistically significant and robust pay-performance link for this group of public sector leaders. While the magnitude of the estimated performance sensitivity is not as large as one sees in the private sector work, the finding is somewhat striking in view of the general assessment that "bureaucrats" ought to be paid flat wage schedules. However, at another level, the finding is not very surprising. Even though the market for head teacher pay is much more regulated than the market for CEOs in a typical private sector setting, all schools operate in a labor market where recruitment and retention is important.

4.3.2 Turnover Estimates

We now turn to the determinants of head teacher turnover. In parallel with the work on private sector CEOs, our model also predicts that we should see

head teachers in poorly performing schools being more likely to leave. To the extent that salaries are downwardly rigid, this may be even more important in public sector settings.

Columns (1) and (2) of Table 4 give the basic empirical findings where the probability of head teacher turnover is related to school performance in the previous year. The results show a strongly significant negative coefficient on lagged school performance, confirming that head teachers are much more likely to move on from less well performing schools. Both the school performance measures based on exam scores are strongly significant and have the expected signs. As reported in column (2), this finding remains the case when we control for the age of teachers. Thus, head teachers in schools with lower academic achievement are more likely to leave. This is also consistent with the salary results above suggesting that salary incentives keep higher paid teachers in place in better performing schools.

We also consider whether there are age related differences in head teacher turnover which could be interpreted as variations in γ in the theoretical model varying with age. We investigate this by estimating turnover models for three age groups: those aged less than 45, those between 45 and 55, and those aged over 55. The results show that turnover increases significantly more with age in the oldest group. Moreover, the impact of having a larger fraction of failing pupils at GCSE level is much larger in this group. However, there is little variation in the effects of the basic exam performance measure based on the proportion getting 5 or more grades A* through C in different age groups. These results are consistent with γ being somewhat higher in the oldest age group (heads aged over 55).

Taken together these results are consistent with the basic prediction of the model – that downward wage flexibility of head teachers makes turnover sensitive to school performance in so far as the latter reveals something about the quality of the head teacher. This finding is consistent with the model in section 2.2 above.

4.4 Robustness

In this section, we discuss a number of additional specifications that we ran in order to assess the robustness of the basic findings.

One possible concern with our results is that, even with school and head teacher fixed effects, we have failed to account fully for match specific utility. An additional way to check for this is to look at the salary paid to a head

teacher when he/she first joins a school so as to determine whether there is any evidence that good schools pay an *ex ante* wage premium. This is investigated in Table 5, columns (1) and (2) which runs a regression of the salary paid to a head teacher when they are first hired by a school on the latest measure of school performance. The coefficient on the main school performance measure is estimated to be negative. Thus, if anything, there is a negative premium to being recruited into a good school.

In column (3) of Table 5, we add the initial salary of a head teacher as an additional regressor in our pay-performance regression. We find that there is no significant correlation between the initial salary and subsequent salary after we condition on school fixed effects, even though we continue to get a significant pay-performance gradient. This makes it less likely that our results are driven by selection.[7] Moreover, this finding suggests that the pay-performance gradient that we have uncovered is primarily for retention rather than recruitment purposes.

Our specification for turnover does take account of heterogeneity in the rules for turnover at the school level, contrary to what we would expect from equation (4). We investigate this issue further in Table 6 where we carry out a strong test and include a set of school fixed effects. As a benchmark, we include column (2) from Table 4 as column (1) in Table 6. In column (2) we show the same specification using a linear probability model, where the results are shown to be both qualitatively and quantitatively similar. We then report the results from running a linear probability model with school fixed effects in column (3). This continues to display a strong association between school performance measures and turnover.

5 Conclusions

This paper has investigated whether public sector CEOs, in the guise of school principals, are paid for performance. Given the focus on the leadership role of principals in improving school quality, it makes theoretical sense to believe that they would. Even though they qualify as bureaucrats in the conventional use of that term, they are recruited and retained in a competitive labor market where there is some flexibility to reward good performance.

[7] It is interesting to note that there is a positive, significant correlation between current salary and initial salary if we exclude the school fixed effect from the regression in column (3).

Thus, even though the pay-performance gradient may be smaller in the context of schools, the logic of the career concerns approach applies equally well in this context. In this sense, when it comes to leadership in schools, the class of public sector CEOs that we have studied are not fundamentally different from the private sector CEOs which have been studied previously.

The central results in the paper relate pay to performance and performance to turnover. In both cases, the relationship that we find is extremely robust. In particular, it survives the inclusion of both school and individual head teacher fixed effects. The analysis has taken advantage of a unique context for studying these issues, which is made possible by being able to match head teacher salaries to the performance of the schools in which they work.

The results point towards the importance of recognizing that labour markets for public sector leaders may play an important role in rewarding performance when organizations have some flexibility to adjust pay for the purposes of recruitment and retention. The debate around public sector reform has frequently focused too quickly on the role of explicit incentive schemes for public servants, forgetting that implicit incentives are already working through the labour market.

There are many further issues that need addressing in the study of public sector leaders and their role in shaping public organizations. While it may be true that public sector leaders are more motivated by mission than money, this does not make money irrelevant, especially in a market context. In the model that we have put forward, we have abstracted entirely from the effort inducing role of performance-pay. In future work, it would be interesting to think of ways to investigate whether the patterns of performance over the career cycle suggest less slacking towards the career end in line with models of intrinsic motivation. It would also be interesting to see whether other aspects of job design such as the autonomy of the head are used as complementary tools for attracting high quality heads to lead schools.

References

[1] Abowd, John and Francis Kramarz, [1999], "The Analysis of Labor Markets Using Matched Employer-Employee Data", in Orley Ashenfelter and David Card (eds), *Handbook of Labor Economics*, Amsterdam: North-Holland.

[2] Bertrand Marianne, and Antoinette Schoar, [2003], "Managing with Style: The Effect of Managers on Firm Policies", *Quarterly Journal of Economics*, 118(4), 1169-1207.

[3] Besley, Timothy and Maitreesh Ghatak [2005], "Competition and Incentives with Motivated Agents", *American Economic Review*, 95(3), 616-636.

[4] Farber, Henry and Robert Gibbons [1996], "Learning and Wage Dynamics", *Quarterly Journal of Economics*, 111(4), 1007-47.

[5] Gibbons, Robert and Kevin J. Murphy, [1992], "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence", *Journal of Political Economy*, 100(3), 468-505.

[6] Hall, Brian and Jeff Leibman, [1998], "Are CEOs really paid like bureaucrats?", *Quarterly Journal of Economics*, 113(3), 653-691.

[7] Harris, Sandra, [2006], *Best Practices of Award-Winning Secondary School Principals*, Corwin Press.

[8] Heckman, James and Alan Krueger, [2004], *Inequality in America: What Role for Human Capital Policies?*, MIT Press

[9] Holmstrom, Bengt, [1999], "Managerial Incentive Problems: A Dynamic Perspective", *Review of Economic Studies*, 66(1), 169-82.

[10] Hoxby, Caroline, [2003], *The Economics of School Choice*, University of Chicago Press.

[11] Jensen, Michael and Kevin Murphy, [1990], "Performance Pay and Top-Management Incentives", *Journal of Political Economy*, 98(2),225-64.

[12] Machin, Stephen, [2008], "The New Economics of Education - Methods, Evidence and Policy", *Journal of Population Economics*, 21(1), 1-19.

[13] Machin, Stephen and Anna Vignoles, [2005], *What's the Good of Education? The Economics of Education in the United Kingdom*, Princeton University Press.

[14] Murphy, Kevin J., [1999], "Executive Compensation," in Orley Ashenfelter and David Card (eds), *Handbook of Labor Economics*, Amsterdam: North-Holland.

[15] Policy Exchange, [2007], *The Leadership Effect*, Policy Exchange Limited.

Table 1: Descriptive Statistics

	1994	1995	1996	1997	1998	1999	2000	2001	2002
A. Unbalanced Panel									
Average salary (£'000)	39.8	41.4	42.6	44.6	46.2	48.2	51.7	55.0	58.1
Proportion Getting 5 or More A*-C GCSE Grades	.39	.39	.41	.41	.42	.44	.45	.46	.48
Proportion Getting No GCSE Passes	.07	.07	.07	.07	.05	.05	.05	.04	.04
Turnover Proportion	-	.08	.10	.10	.13	.09	.10	.13	.12
Number of Schools	2478	2506	2533	2518	2497	2469	2471	2314	2227
B. Balanced Panel									
Average salary (£'000)	39.9	41.5	42.8	44.8	46.4	48.4	51.8	55.2	58.2
Proportion Getting 5 or More A*-C GCSE Grades	.39	.40	.41	.42	.43	.45	.45	.47	.48
Proportion Getting No GCSE Passes	.06	.07	.06	.06	.05	.04	.04	.04	.04
Turnover Proportion	-	.08	.10	.09	.12	.09	.09	.12	.11
Number of Schools	1790	1790	1790	1790	1790	1790	1790	1790	1790

Notes: Salary data (in £'000s) from Database of Teacher Records; Proportion Getting 5 or More A*-C GCSE (General Certificate of Secondary Education) Grades and no GCSE Passes from School Performance Tables; Turnovers calculated from Database of Teacher Records. Overall unbalanced panel sample size is 22013, in 2601 schools with at least four continuous time series observations and 4309 head teachers between 1994 and 2002. The balance is: 4 observations – 30 schools; 5 observations – 39 schools; 6 observations – 69 schools; 7 observations – 210 schools; 8 observations – 463 schools; 9 observations 1790 schools.

Table 2: Comparison of Explanatory Power of Teacher and School Fixed Effects in Head Teacher Salary Equations, 1994-2002

Unbalanced Panel
(2601 Schools, 4309 Head Teachers; Sample Size = 22013)

	(1)	(2)	(3)	(4)	(5)	(6)
Head Teacher Fixed Effects	F(4309, 17704) = 3.97 [p = .00]	No	F(4309, 15103) = 1.42 [p = .00]	F(4309, 17696) = 37.38 [p = .00]	No	F(4309, 15095) = 6.28 [p = .00]
School Fixed Effects	No	F(2601, 19412) = 5.29 [p = .00]	F(2601, 15103) = 1.29 [p = .00]	No	F(2601, 19404) = 29.49 [p = .00]	F(2601, 15095) = 2.14 [p = .00]
Year Dummies	No	No	No	Yes	Yes	Yes
R-Squared	.491	.415	.584	.950	.900	.964

Balanced Panel
(1790 Schools, 3033 Head Teachers; Sample Size = 16110)

	(1)	(2)	(3)	(4)	(5)	(6)
Head Teacher Fixed Effects	F(3033, 13077) = 4.11 [p = .00]	No	F(4309, 11286) = 1.28 [p = .00]	F(3033, 13068) = 46.72 [p = .00]	No	F(3033, 11278) = 6.31 [p = .00]
School Fixed Effects	No	F(1790, 14320) = 5.26 [p = .00]	F(2601, 15103) = .88 [p = 1.00]	No	F(1790, 14311) = 35.69 [p = .00]	F(1790, 11278) = 1.54 [p = .00]
Year Dummies	No	No	No	Yes	Yes	Yes
R-Squared	.488	.397	.551	.959	.911	.967

Notes: The dependent variable is log(salary).

Table 3: Pay-Performance Regressions For Secondary School Head Teachers, 1994-2002

	Unbalanced Panel (2601 Schools, 4309 Head Teachers; Sample Size = 22013)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Proportion Getting 5 or More A*-C GCSE Grades	.113 (.009)	.057 (.009)	.038 (.009)	.032 (.009)	.040 (.007)	.028 (.005)	.022 (.005)
Log(Number of Pupils in School)			.141 (.008)	.143 (.008)		.145 (.004)	.146 (.008)
Proportion Getting No GCSE Passes				-.066 (.016)			-.054 (.013)
Age and Gender Controls	Yes	Yes	Yes	Yes	No	No	No
School Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Head Teacher Fixed Effects	No	No	No	No	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	.549	.913	.917	.917	.964	.966	.966

	Balanced Panel (1790 Schools, 3033 Head Teachers; Sample Size = 16110)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Proportion Getting 5 or More A*-C GCSE Grades	.105 (.011)	.059 (.010)	.042 (.010)	.036 (.010)	.044 (.009)	.032 (.008)	.028 (.006)
Log(Number of Pupils in School)			.135 (.010)	.137 (.009)		.144 (.007)	.145 (.005)
Proportion Getting No GCSE Passes				-.059 (.018)			-.047 (.013)
Age and Gender Controls	Yes	Yes	Yes	Yes	No	No	No
School Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Head Teacher Fixed Effects	No	No	No	No	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-Squared	.562	.926	.929	.929	.967	.970	.970

Notes: The dependent variable is log(salary). Standard errors (clustered on head teacher id) in parentheses for specifications (1) to (4) excluding head teacher fixed effects. Bootstrapped standard errors for specifications (5), (6) and (7) including head teacher fixed effects.

Table 4: Head Teacher Turnovers

Unbalanced Panel
(2601 Schools, 4130 Head Teachers; Sample Size = 19287)

	(1)	(2)	(3)	(4)	(5)
	All	All	Age<45	Age≥45 and Age<55	Age≥55
Proportion Getting 5 or More A*-C GCSE Grades (t-1)	-.065	-.102	-.143	-.099	-.123
	(.014)	(.015)	(.037)	(.017)	(.051)
Proportion Getting No GCSE Passes (t-1)	.175	.138	.080	.048	.485
	(.049)	(.049)	(.097)	(.052)	(.184)
Age of Head Teacher		.011	.005	.003	.036
		(.001)	(.003)	(.001)	(.003)
Male Head Teacher		-.010	-.010	-.009	.013
		(.006)	(.011)	(.006)	(.021)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Sample Size	19287	19287	2202	12785	4300

Balanced Panel
(1790 Schools, 2919 Head Teachers; Sample Size = 14320)

	(1)	(2)	(3)	(4)	(5)
	All	All	Age<45	Age≥45 and Age<55	Age≥55
Proportion Getting 5 or More A*-C GCSE Grades (t-1)	-.084	-.124	-.113	-.120	-.176
	(.017)	(.018)	(.041)	(.019)	(.061)
Proportion Getting No GCSE Passes (t-1)	.120	.063	.115	-.030	.261
	(.060)	(.060)	(.112)	(.063)	(.228)
Age of Head Teacher		.011	.006	.003	.040
		(.001)	(.004)	(.001)	(.004)
Male Head Teacher	-.010	-.011	-.014	-.008	.005
	(.006)	(.007)	(.014)	(.007)	(.024)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Sample Size	14320	14320	1570	9472	3278

Notes: The dependent variable is a binary variable indicating head teacher turnover between (t-1) and t. Marginal effects from probit models; standard errors (clustered on head teacher id) in parentheses.

Table 5: New Head Teacher (First Year in Post) Salary Equations and Estimated Panel Models For Sample of New Entrants

	Unbalanced Panel (1925 Head Teachers in 2074 Schools)		
	Initial Salary in Entry Year		Panel Model (Model (4) in Table 3 plus Initial Salary)
	(1)	(2)	(3)
Proportion Getting 5 or More A*-C GCSE Grades	-.048	-.045	.033
	(.016)	(.016)	(.016)
Log(Number of Pupils in School)	.206	.204	.144
	(.005)	(.005)	(.016)
Proportion Getting No GCSE Passes	.050	.039	-.063
	(.054)	(.054)	(.029)
Initial Entry Year Log(Salary)			.012
			(.023)
Age and Gender Controls	No	Yes	Yes
Year Dummies	Yes	Yes	Yes
School Fixed Effects	No	No	Yes
R-Squared	.697	.703	.922
Sample Size	2074	2074	8229

	Balanced Panel (1371 Head Teachers in 1435 Schools)		
	Initial Salary in Entry Year		Panel Model (Model (4) in Table 3 plus Initial Salary))
	(1)	(2)	(3)
Proportion Getting 5 or More A*-C GCSE Grades	-.060	-.058	.028
	(.020)	(.020)	(.017)
Log(Number of Pupils in School)	.213	.212	.146
	(.005)	(.005)	(.021)
Proportion Getting No GCSE Passes	.003	.003	-.056
	(.059)	(.059)	(.031)
Initial Entry Year Log(Salary)			.016
			(.031)
Age and Gender Controls	No	Yes	Yes
Year Dummies	Yes	Yes	No
School Fixed Effects	No	No	Yes
R-Squared	.742	.745	.929
Sample Size	1435	1435	5838

Notes: The dependent variable is log(salary). Standard errors (clustered on head teacher id) in parentheses.

Table 6: Head Teacher Turnovers – Including School Fixed Effects

	Unbalanced Panel (2601 Schools, 4130 Head Teachers; Sample Size = 19287)		
	(1)	(2)	(3)
	Model (2) From Table 4 – Probit Marginal	Linear Probability Model	Linear Probability Model With School Fixed Effects
Proportion Getting 5 or More A*-C GCSE Grades (t-1)	-.102	-.093	-.173
	(.015)	(.015)	(.048)
Proportion Getting No GCSE Passes (t-1)	.138	.175	.073
	(.049)	(.057)	(.088)
Age of Head Teacher	.011	.012	.027
	(.001)	(.001)	(.001)
Male Head Teacher	-.010	-.009	-.008
	(.006)	(.006)	(.021)
Year Dummies	Yes	Yes	Yes
School Fixed Effects	No	No	Yes
Sample Size	19287	19287	19287

	Balanced Panel (1790 Schools, 2919 Head Teachers; Sample Size = 14320)		
	(1)	(2)	(3)
	Model (2) From Table 4 – Probit Marginal	Linear Probability Model	Linear Probability Model With School Fixed Effects
Proportion Getting 5 or More A*-C GCSE Grades (t-1)	-.124	-.116	-.143
	(.018)	(.018)	(.054)
Proportion Getting No GCSE Passes (t-1)	.063	.093	.013
	(.060)	(.070)	(.102)
Age of Head Teacher	.011	.012	.027
	(.001)	(.001)	(.001)
Male Head Teacher	-.011	-.011	-.016
	(.007)	(.007)	(.024)
Year Dummies	Yes	Yes	Yes
School Fixed Effects	No	No	Yes
Sample Size	14320	14320	14320

Notes: The dependent variable is a binary variable indicating head teacher turnover between (t-1) and t. Marginal effects from probit models; standard errors (clustered on head teacher id) in parentheses.