CMPO Working Paper Series No. 01/36

The Impact of Teacher Pay Reform on Pupil Attainment: An outline of the CMPO research project on the Performance Threshold

Simon Burgess[1]
and
Bronwyn Croxson[2]

[1]CMPO, University of Bristol and CEPR
[2]CMPO, University of Bristol

August 2001

Abstract
In 2000 the government introduced a nationwide reform of teacher pay in England, which linked increases in salary to the performance of individual teachers. This paper outlines a research project which evaluates the impact of the new pay regime on pupil attainment. The paper provides a brief summary of the nature of the new pay regime and relevant aspects of the reforms, an outline of the method and econometric strategy, a description of the various sources of data and the strategy for obtaining data from schools, and an outline of progress to date in obtaining this data.

JEL Classification: J33, J44, J45
Keywords: performance related pay, teacher effectiveness

Acknowledgements
All of the ideas and work described in this paper have been jointly produced with Adele Atkinson, Paul Gregg and Carol Propper, all at CMPO. We thank the Leverhulme Trust for funding CMPO and this project. We are also deeply grateful to all the schools that are taking part in the research, and without whose assistance and effort none of the work would be possible. We also thank officials at the Department for Education and Skills (Department for Education and Employment during the initial stages of the project) for advice, relevant commentary, and access to national datasets. Finally, we thank David Reynolds, Anna Vignoles and Ted Wragg and his team at Exeter University for advice.

Address for Correspondence
CMPO, Department of Economics
University of Bristol
8 Woodland Road
Bristol
BS8 1TN
Tel: +44 (0)117 928 8436
Simon.burgess@bristol.ac.uk

CMPO is funded by the Leverhulme Trust.

1. Introduction

In 2000 the government introduced a nationwide reform of teacher pay in England, which linked increases in salary to the performance of individual teachers. The new regime, known as the "Performance Threshold", is supposed to reward teachers' achievements in a number of dimensions, including their impact on improved pupil test and exam results.

This paper outlines a research project which evaluates the impact of the new pay regime on pupil attainment. The project forms part of a programme of work analysing Incentives in Public Sector Organisations, in the Leverhulme Centre for Market and Public Organisation (CMPO). CMPO is funded by the Leverhulme Trust to conduct research about the changing boundary between the state and private sectors, including the effect of introducing explicit incentives into public sector agencies. The very small body of directly relevant empirical evidence which exists in this area suggests that in the public sector explicit incentives may have unintended, undesirable, consequences. [Burgess and Metcalfe 1999] The project described in this paper will contribute evidence relevant both to the general issue of incentives in the public sector and to education policy.

With the exception of a small number of surveys of teacher attitudes [Marsden 2000; Purslow 2000], there is no UK empirical evidence directly relevant to the likely impact of the Performance Threshold. There is, however, a growing body of commentary. [Richardson 1999a,b; Richardson 2000; Storey 2000; Thompson 2000] Commentators generally back up their arguments with qualitative evidence relating to the introduction of performance related pay in other parts of the public sector or in US schools. As outlined in an accompanying literature review, there is actually rather little quantitative evidence about the impact on teacher behaviour of performance related pay. [Burgess, Croxson, Gregg and Propper 2001]

The research project reported in this paper will quantitatively analyse the impact of the Performance Threshold on pupil attainment. In the project we assess whether the possibility of crossing the Performance Threshold and thereby increasing their salary leads individual teachers to exert more effort to improve pupils' exam scores. We use

data on individual pupils' test scores, with pupils linked to individual teachers, and data on the eligibility of their teachers. The nature of the UK education system makes it possible to use "value-added" scores. We compare the "value-added" scores associated with each teacher before and after the introduction of the Threshold. The nature of the scheme means that we can also compare any change in value-added scores across teachers, comparing those who are eligible to pass the Threshold with those who are not. Information obtained from national datasets will be used to control for other factors likely to affect pupil performance, including school characteristics, teacher characteristics, general changes over time, and pupil characteristics. This project is being conducted over three years, with the first stage of the data collection occurring in 2000-01 (collecting data covering 1997 – 1999), and the second stage occurring in 2002-03 (covering 2000 – 2002).

Parallel to the quantitative project reported in this paper, a qualitative study of 25 Headteachers' experience with the first Threshold application process and of incentives in schools has been conducted. [Croxson 2001] Other research studies are underway at the University of Exeter and at the London School of Economics. These studies are complementary, since they are qualitative, and investigate the impact of the Threshold on dimensions other than pupil attainment.

This paper is designed to disseminate information about the methodology, and about some of the issues arising during data collection. The paper provides a brief summary of the nature of the new pay regime and relevant aspects of the reforms, an outline of the method and econometric strategy, a description of the various sources of data and the strategy for obtaining data from schools, and an outline of progress to date in obtaining this data.

2. The nature of the Performance Threshold pay scheme for teachers

The reforms to teacher pay

During the second half of the nineteenth century teachers in English state secondary schools were paid according to students' exam results, but this was abandoned because it was believed to reward teachers who concentrated on more able

pupils.[Hood et al. 1999] Formal systems of performance related pay were not used in state schools during most of the twentieth century. Interest in performance related pay grew during the 1980s, stimulated by a perception that teaching standards were poor and contributing to low educational attainment and, perhaps, to poor economic performance. [Tomlinson 1992, 2000] During the 1990s, successive administrations attempted to introduce some form of performance related pay into state schools, but succeeded only in introducing performance related pay for head teachers and their deputies. [Marsden and French 1998]

The 1997 Labour administration signalled a range of reforms to education, in a Green Paper published in 1998: "Teachers: Meeting the Challenge of Change". The reforms included the introduction of a performance-related element in teacher pay, the "Performance Threshold", designed to affect teacher effort as well as recruitment into and retention within the profession. The Green Paper argued that teacher motivation was adversely affected by a "culture" which did not recognise and reward outstanding performance.

Prior to the introduction of the new scheme, all teachers were paid on a unified basic salary scale, with nine full points, ranging from £14,658 - £23,193 p.a. (2000 prices). An individual's position on the scale depended on his/her qualifications and experience, and teachers usually progressed up the scale in annual increments. In addition to the basic salary, they might be in receipt of management, excellence, or recruitment and retention points.[School Teachers' Review Body 2000] About 75% of teachers were at the top of this scale, at spine point 9.

After the reforms, teachers at spine point 9 could apply to pass the Performance Threshold. Passing the Threshold gives teachers an annual bonus of £2000, payable without revision until the end of their career and able to be included in calculations of pensionable salary. It is therefore of significant lifetime value. Once over the Threshold, teachers move onto a new "Upper" pay scale. Details of the upper scale have not been confirmed, but it has been announced that it is likely to comprise an additional four points up to an annual salary of £30,000, with increments explicitly related to performance. Threshold payments are currently funded out of a separate

budget, administered by central government, with no limit or quota on the number of teachers allowed to obtain Threshold payments.

In order to pass the Performance Threshold, individual teachers have to demonstrate that they have reached acceptable standards in five areas: knowledge and understanding of teaching; teaching management and assessment; wider professional effectiveness; professional characteristics; and pupil progress. [DFEE 2000] In the area of pupil progress, the focus of this project, the Threshold application form gave teachers the following instruction:

> *Please summarise evidence that as a result of your teaching your pupils achieve well relative to their prior attainment, making progress as good or better than similar pupils nationally. This should be shown in marks or grades in any relevant national tests or examinations, or school based assessment for pupils where national tests and examinations are not taken.* [DFEE 2000]

Teachers were required to complete application forms by July 2000, writing what were in effect short essays demonstrating their performance in each area. 80% of eligible teachers applied. Head Teachers then assessed applications and recommended whether or not individual teachers should pass the Threshold. Each school was then to be audited by an external assessor. Teachers who are not passed have limited rights of appeal. A recently published study [Wragg et al. 2001], conducted a survey of a random sample of 1000 schools in over 150 local education authorities. They found that in these schools, 88 percent of the eligible teachers applied, and of these 97 percent were awarded the bonus. Payments will eventually be backdated to September 2000.

Commentary and evidence on recent reforms

Many commentators and all of the teacher unions are critical of the Performance Threshold. Commentators argue that it is unlikely to motivate teachers since there is some evidence that public sector workers in general, and teachers in particular, rate intrinsic rewards more highly than extrinsic rewards. [Richardson 1999a; Thompson 2000] It is also argued that any impact on effort is diluted by the large number of

goals teachers are set in the Threshold process, that the subjectivity of the assessment process undermines any link between effort and reward, and that teachers are being assessed on dimensions not fully under their control. [Richardson 1999b, 2000] Storey analysed submissions made during the consultation process following publication of the Green Paper, and found that most submissions argued that the proposals would increase bureaucracy, undermine team work in schools, and that the available measures of pupil attainment were unreliable. Many submissions also argued that teaching quality cannot be measured, rejecting the principle of performance related pay for teachers. [Storey 2000]

To date the results of two surveys of teacher attitudes to the Performance Threshold have been reported. The first was conducted before teachers applied for the Threshold. [Marsden 2000] The majority of respondents disagreed with the principle of performance related pay, most believed that £2000 was insufficient to induce extra effort, and just over half disagreed with principle of linking pay to pupil exam results. The second survey was conducted after teachers had applied for Threshold payments, but before the start of the assessment process. [Purslow 2000] A substantial number of respondents were concerned that the process would not be fair, that it would be divisive, and that the standards did not accurately describe good teaching. The survey did not investigate their attitude to using pupil attainment as a standard, but did ask whether respondents had found it straightforward to complete the Threshold application form. About a quarter experienced difficulty completing the section on pupil progress: 43% had difficulty getting relevant statistics; 23% said there was no formal testing of their pupils; and 10% expressed reservations about quantifying pupil progress.

As noted, a recent study [Wragg et al. 2001] found that in their sample of schools, 88 percent of the eligible teachers applied, and of these 97 percent were awarded the additional payment. Unsuccessful candidates were deemed to be failing on one or more aspect of their teaching. Head teachers in the study reported that they did not find it difficult to assess the five standards that teachers had to meet to receive their £2000 performance payment (which included pupil performance), but that the process was extremely time consuming. Heads were generally in agreement with the external assessors who were responsible for verifying the outcomes (granting of the award),

though they felt this verification procedure was extremely time consuming and costly. Around 60 percent of heads indicated that they were against PRP, in principle, while 39 percent were in favour in principle. However, within this latter group, a large majority expressed concerns about its current implementation.

In terms of impact on teachers' actions in the classroom, three quarters of heads felt that the assessment had made little or no difference to what teachers did. Preliminary evidence from a parallel study by the same team, which looked at classroom behaviour during the year assessment period, suggests that teachers might have improved their recording methods, rather than change the way they teach.

3. Our Research Approach

Our prime research question is whether the introduction of the Performance Threshold for teachers resulted in any change in pupil test score outcomes. We are primarily interested in teachers and the pupils they teach. We look for differences between teachers eligible for the Threshold and those ineligible, before and after the reform. This research question drives our data acquisition strategy and our econometric approach. These are discussed in the next two sections, but here we set out the broad outlines of our approach.

To provide convincing evidence, we need to be able to link teachers to the actual pupils they taught for specific subjects, not just pupils they might have taught as they belong to the same school. The data link between pupils and teachers is simply the class list – the list of pupil names associated with a teacher for a specific period of time for a specific subject. This information is held **only** in schools. This immediately determines one of the key features of our study – we need necessarily to collect primary data from within schools, as only they hold the crucial linking data. Analysis of secondary data sources (for example national datasets) can only approximate this linkage.

When analyzing the test score outcomes associated with a teacher, we need to control for the initial skills of the pupils s/he taught. Therefore, we need to use value-added data – that is, test score *gains*. The best opportunity that the UK education assessment

and information system provides is between the Key Stage 3 exams at age 14 and the GCSE exams at age 16. This therefore is the age range we study. Key Stage exams are taken in english, maths and science; pupils also have to take GCSE exams in these subjects (among others). Using value-added scores in all these different exams gives us three observations per pupil and also within-pupil variation.

We also need to control for the teacher's ability. Our question is about whether the Performance Threshold motivated teachers to supply more effort. We cannot allow this effect to be confounded with the fact that more able teachers are more likely to get better results and are also more likely to receive the pay bonus. Therefore, we need to look at the *same* teachers both before and after the introduction of the reform. This forces us to undertake a longitudinal study, maintaining a relationship with schools for two complete teaching cycles – that is, a two-year (KS3 → GCSE) cycle before the Threshold and another after. This also obviously means that we are necessarily not looking at the same *pupils*. We control for pupil effects by the use of value-added data and other approaches (see below).

Having acquired class lists from schools, we could in principle get exam scores for the pupils from national datasets held by DfEE. However, we were advised that the matching of these between school and DfEE records might be unreliable as at that time there were no unique national pupil identifiers. Since the schools themselves held the data, we decided to collect the scores data directly. Similarly, while information on the teachers themselves was also available in principle from national data (the Database of Teacher Records), problems of confidentiality meant that it would be difficult to map this into particular teachers in our schools. Again, since we were collecting data from schools, we decided to collect this information locally too.

All this data allows us to control for pupil effects, class effects, teacher effects, subject effects and school effects. Using this, we hope to determine in a convincing way whether or not the introduction of the Performance Threshold has had any impact on pupil achievement.

4. Data

Approaching schools

In order to link pupil attainment with individual teachers, we need to use data on class lists, which is held only in individual schools. To this end we have approached schools directly, inviting them to participate in the study.

Our objective is to have a sample of at least 60 participating schools in order to have 500 teachers in the sample. Early results suggested that we may get, on average, ten English Maths or Science teachers per school who have been in the school over the entire relevant period (1997/8-2001/2). We anticipate that some schools will drop out before the second stage of data collection.

Schools have been approached in five tranches. The first tranche comprised a pilot group of 40 schools, chosen from a group of about 200 schools who had previously participated in a DfEE pilot value-added exercise. These schools were therefore known to be research aware and to have good information systems. We selected the 40 schools as a convenience sample of the schools closest to Bristol. The second tranche comprised the remaining schools from the DFEE pilot group. Third and subsequent tranches comprised schools in particular groups of LEAs, chosen to reflect a balance of school types and urban/rural features. The criteria for selecting schools in each tranche are summarised in Table 1.

Schools are approached by being sent a covering letter inviting them to participate, as well as a project outline and a stamped-addressed postcard to return indicating whether they are willing to participate, are not willing to participate, or want more information before taking a decision. Their addresses are obtained from the DFEE Performance Tables. The approach is personalised by addressing letters to named Head Teachers. The names of the first tranche of heads were obtained from a published DFEE source [The Education Authorities Directory and Annual 1999]. This proved to be out of date, so the names of heads in subsequent tranches have been obtained by telephoning each school.

The first tranche of schools were all telephoned after the initial, letter-based approach. Schools in the first and second tranches who did not respond to the initial approach were sent a follow-up letter and subsequently telephoned. The phone calls were extremely time consuming (up to eight calls had to be made to each school to speak to the head, or to draw the issue to the head's attention), and only a small number of schools agreed to participate as a result. (By November 1st 17 of the 54 schools who responded to the initial approach said yes, compared with 5 of the 53 schools who were followed up by telephone. These numbers do not tally with those shown in Table 1 because a number of these schools subsequently withdrew from the study.) Intensive follow-up was therefore dropped from subsequent tranches, in favour of more extensive sampling.

After agreeing to participate, schools are sent a letter guaranteeing the confidentiality and security of the data, a schedule of the data we need and, to obtain supplementary information, are either interviewed or sent a questionnaire. The data we request from each participating school is summarised in Table 2. Data received from schools are stored anonymously in a completely secure environment; upon receipt, we ensure that the data has been anonymised fully.

The response from schools

As shown in Table 1, by May 2001 48 schools have agreed to take part in the study. We have to date received data from 15 schools. As shown in Table 1, a large number of schools have not responded, or have indicated that they do not wish to take part. The overall acceptance rate has been 3%. Some schools have volunteered information about why they do not wish to take part, either over the phone or in a letter. Their responses have been categorised and are shown in Table 3. The most frequently cited reason for not taking part has been lack of time, a category which includes feeling under pressure from the large number of initiatives directed to schools. As stated by one headteacher:

> *Worthy and interesting as your research sounds, this school has been so bludgeoned and battered by bureaucracy and data collection – apparently "at*

random" from every acronym under the sun – HMI, LEA, Ofsted, Dfee, EiC, NFER, QCA, NAO, EBP, LLLP ad infinitum – that we can't take any more.

A number felt they were involved in too many research projects already. A small number stated that they were not willing to participate because they disagreed with the Performance Threshold in principle or did not believe it would affect (or that it was designed to affect) pupil attainment. In addition, a small number stated directly that they did not wish to take part because they disagreed with our research design.

Some schools stated that they could not take part because they did not have the required data. These schools usually lacked relevant class lists. In addition to those shown here, a number of schools who initially agreed to participate have subsequently had to decline, usually because they discover that they do not have the requisite data. A small number of schools have withdrawn because they have decided that it would be too time consuming to collect the data.

We have agreed to pay schools if they have requested it, and sometimes offered payment to schools who have said that they have the data but do not have the time to collate it. We have offered a maximum of £60 per school and, in the few instances (six) where it has been accepted, have paid only on receipt of the data. Two schools wanted payment but believed £60 to be insufficient (it was described by one as "*peanuts, in fact its only one peanut"*). We have also offered schools the services of a research assistant to download or collate information in the school. No school has needed this help. In one instance the researcher visited a school to help collate data, but found that the school had already done so.

Of the schools who have accepted, to date data has been received in some form from only 15, and complete datasets from 11. This reflects the substantial pressure facing heads and their deputies, who are helpful and apologetic, but extremely busy. Different schools have involved different members of staff in data collection. In some schools data has been compiled by the heads themselves, and in others by the head teacher's secretary, general administrative staff, or teachers responsible for data management.

Reasons given by head teachers for taking part in the study

Since potential selection problems are an important issue for this project, we collected information on why schools chose to participate; this complements a statistical analysis of selection. As noted above, 25 head teachers were interviewed about their views on the Performance Threshold and on incentives in schools. (The results of these interviews will be reported elsewhere.) This section reports their responses when asked an open-ended question about why they participated in the study. We asked this question to get more information about conducting this type of project, to try and increase the response rate, and as part of the process of determining obvious bias in participating schools. Their responses to this question have been analysed using content analysis, in NVIVO.

Two general categories emerged: schools which are participating because the head believes research to be important; and schools participating because the head is interested in the effect of the Performance Threshold.

Most of those who said they were participating because they believe in the importance of research said that they felt it was their duty to do so. As expressed by one head: "… if I want people to provide research data that's going to help me run the school I've got some kind of duty to participate when people are asking me to help with it." (School 199) One said that he/she was participating because it is important that people gain access to schools, since education-research should be conducted in schools not in isolation. In one case, a head who is him/her self research-active said that he/she is taking part because they know from experience how difficult it is to get schools to participate in research. This head was one of a small number who said that they are, themselves, research active, perhaps implying that they felt some form of empathy with us as fellow researchers.

Two heads stated that their interest in research reflects their efforts to try and ensure that any new initiatives within the school are evidence-based, and to encourage the use of research outputs to improve teaching. Another said that he/she is trying to promote research-awareness since the school has traditionally been "enclosed" and he/she wants to encourage its teachers to be more outwards looking. This head also

stated that policy should be research-based, and that research will be policy-relevant only if schools are directly involved. Another head similarly stated that he/she wished to help research that might influence policy, "in order that perhaps central government might listen a little bit to the effect that this kind of thing has on schools." (School 29)

One head teacher said that he/she believes that his/her school has a number of particularly interesting features, meaning that it should be included in research projects.

A small number of heads said that they were taking part because they believed that the results would be directly useful to them. Four said that they would like to be able to see and use the results relating to their school. One of these heads is an external assessor, and said that the results were also likely to be directly useful to him/her "professionally". Another said that he/she was very interested in finding ways of improving education standards, but that the school is under "severe pressure" with high levels of pupil deprivation and of staff turnover. This head said that "there are some pretty dramatic - worrying factors around so any research that might throw some light on any of that has got to be a good idea" (School 114)

A number of heads cited factors directly related to the Performance Threshold as the reason they are taking part. Three had no firm views about whether the Threshold is appropriate, but nonetheless expressed an interest in finding out whether there is a link between the threshold and pupil attainment or teacher behaviour. One of these said that if there is indeed a link, it should be exploited to improve outcomes and that schools could find this out only by participating in research. Another said that he/she was taking part out of "natural nosiness" about whether the Threshold will "work".

Some heads said that they were participating because they did not think that the Threshold would motivate teachers. As one said: "I suppose I think it's important that we know whether the huge resources going into this is actually making any difference to pupil performance. Because if it doesn't it's a lot of waste of money. And you know I'm particularly keen that we find that out". (School 7) (The resources to which this head referred included his/her time.) Another head said that she/he was taking part in the study because she/he hoped that its results would support her/his view: " in

a way I don't want you to find a link with Performance Threshold if it's to do with money... But if it's to do with professionalism, professional development and improving practice then I'm very happy." (School 39) One head teacher had consulted the staff, who she/he said were critical of the Threshold and who therefore wished to participate in the research because it would provide relevant evidence.

By contrast, some heads said they were participating because they viewed the Threshold favourably. One stated: "well the Threshold was quite an interesting process. I quite enjoyed doing it and felt I got something out of it and felt quite positive about it. So I thought I may actually be contributing towards a balanced picture within a teaching profession that had largely hated it all. So I thought I might be providing a little bit of balance, maybe." (School 67) One head who said that she/he wished to participate so that the results could be fed back into the school also said that she felt able to participate because the process had not been "traumatic" for the school.

A small number of schools said that they are willing to participate in this particular research project because of Bristol University's good reputation (two schools), or because it is independent of the DFEE (three schools). One of these schools said that independence was important to ensuring that the research is objective, another that they have a low opinion of the DFEE and put most material from them "in the bin".

Sample characteristics

The characteristics of the schools in the sample will be analysed in relation to the national population of schools once the first round of data acquisition is completed.

Data from national datasets

In addition to the data from schools, some data about each of these schools can be derived from national datasets. We have access to the five national datasets, shown in Table 4. In most cases we have access to these data for each year since 1995. From these data we can derive school-level information about teacher and pupil

characteristics, and about the institutional structure of the school. The school-level variables we will be able to construct from these datasets are summarised in Table 5.

5. Method: Econometric Approach

Using data from secondary schools, the CMPO project will use a difference-in-difference estimator to compare the attainment of pupils taught by the same teachers before and after the introduction of the Threshold, some of these teachers being eligible for Threshold payments and some not.

This section sets out the econometric approach we will follow in analyzing these data. Obviously, the final details of the statistical procedures we implement will depend on the characteristics of the data eventually available: the amount of usable data, the correlation structures and so on.

Our primary research question is whether the introduction of the Performance Threshold has had any impact on pupil test scores[1]. In general, this, like other performance-related pay schemes, has two potential effects on output. These are through motivation – current teachers working harder or better – and sorting – higher quality people becoming teachers, attracted by the higher reward to ability. This project addresses only the first of these, as we condition on the attributes of particular teachers. The second component is of interest too, but is not considered here.

We first set out a model of the process we assume to be generating pupil test scores. Economists refer to this as the 'education production function', in analogy with output being produced from labour and capital. Then we discuss our analysis of this model; we will use both non-parametric and parametric approaches. Finally, we discuss identification and selection issues.

Pupil attainment will be measured by "Value Added", in other words the change over time in an individual pupil's exam score. It is widely recognised that, for the purposes of analysing teacher performance, pupil attainment should be measured using some

[1] Clearly, many other research questions can be addressed with the rich data we are assembling.

form of value-added measure. [DFEE 1999; Meyer 1996; O'Donoghue et al. 1997] We will calculate value added by linking individual pupil's results in Key Stage 3 tests with results in GCSE exams; this covers the age range 14 to 16. We will consider pupil attainment scores in the three subject areas covered in Key Stage 3: English, Maths and Science. Within each of these areas pupils can sit a range of papers at GCSE, and we will analyse value added between the basic Key Stage 3 test score and each GCSE paper in the same subject area.

The education production function

Economists have modelled the acquisition of knowledge in a number of ways. [Todd and Wolpin 2001] They discuss the role of current and past school, personal and family inputs. In this project, we have a great deal of information on the school, teacher, and class, some information on the pupil and none on the individual pupil's family[2]. We divide our discussion of these inputs into the following sets of factors:

- School factors
- Class and group factors
- Subject factors
- Teacher factors
- Pupil factors

We discuss these in turn. Obviously, our main focus of interest is in teacher factors. There is no reason to believe that the education production technology is linear or separable, so we will allow for interactions between these factors. For example, some factors may be more important in some subjects; some teachers may be better for low ability pupils etc.. The constraint on modelling this heterogeneity is the eventual size of the dataset.

School Factors

This is often a main focus of attention, but since our data here are a lot more disaggregated, overall school characteristics are essentially background factors. This

may include school resources[3], school 'ethos', and so on. We deal with these in the first instance by school fixed effects.

In a supplementary analysis, we can take the estimates of school fixed effects and relate them to the school characteristics that we do have data on. One such set of variables that we can construct that previous researchers have not had access to (in the UK at least) is a detailed description of the characteristics of the teaching staff as a whole. This may well be related to the progress of an individual pupil in a particular subject, as the acquisition of knowledge is to some extent a joint product (causality, of course, would be a lot harder to determine).

Class Factors

Our data also allow us to investigate the importance of the individual's class or subject group. We can investigate much more deeply than the perennial question of the importance of class size, as we know the prior test score of every student in the class. We can therefore also test for peer group effects, and for more complex effects that are a mixture of class size and peer group. For example, an individual student's progress might depend on how many students in her class have lower test scores than the focus student, how many have higher scores, or the variability of scores and the composition of the group. Our data allow us to investigate these factors in detail.

This issue relates to schools' setting or streaming policies. Schools allocate pupils to sets or groups for individual subjects, and typically do so in accordance with a particular policy (as opposed to randomly). As noted above, we can condition directly on the characteristics of the class, but we have collected information on schools' setting policies. Since we will only have a relatively small number of schools with complete data, it is unlikely that we would be able to establish with any confidence a link between setting policy and student performance.

[2] Given the school's address, we can, however, match in details of the local areas.

[3] It is not clear that one should condition on school resources in any case. Todd and Wolpin point out that to do so risks "including things twice" and causes problems for interpretation. For example, what is the interpretation of 'class size' when school resources are included?

Subject Factors

We can allow features of the knowledge acquisition process to differ between subjects: maths, english and science. Progress in different subjects may be faster at different ages. Also, whilst the Key Stage exams are national, the GCSE exams taken at 16 are not. Schools can choose their exam board, and different schools take different exam boards for different subjects. Different exams may favour different types of student – for example, some may be more suitable for high flyers or others for lower ability students. We can control for exam subject (the unit of observation for some of our regressions is a student-subject-year) and for exam board.

Teacher Factors

The key teacher factors are the ability and effort of the teacher; this is the focus of our main research question. Again, recall that we are considering changes in the effort of incumbent teachers, not changes in the ability distribution in the intake of new teachers.

This analysis is based on fixed teacher effects. As has been described, we follow teachers over two teaching cycles. We do in fact know a good deal about teachers from the DTR dataset, but since this is anonymised, we cannot match this at teacher level to the estimated fixed effects (we can match at school level).

We assume that teacher ability is given by an idiosyncratic fixed effect plus a common ability-experience profile. A teacher's professional expertise is likely to change with teaching experience. People learn from doing the job, and are better able to respond to situations and pupils. Alternatively, disillusionment may set in and effectiveness decrease after a prolonged period in the job. The idiosyncratic component is absorbed into the teacher fixed effect and the experience profile is estimated. As we noted above, the spine point that the teacher is on – which determines eligibility – is important for identification and is closely related to experience, so the estimating the profile is important (see below).

We now turn to teacher effort, which is unobserved. 'Effort' here summarises the amount of care, interest, concern, efficiency, and imagination that teachers supply per

hour, as well as simply the number of (uncontracted) hours worked. Economic analysis suggests that teacher effort depends on the marginal wage (the marginal return to effort) and an idiosyncratic fixed effect. The fixed effect captures inherent differences in the desire to supply effort[4]. We can also allow for a number of specific factors, that may influence the teacher's propensity to exert effort. These are the general school environment, and the average ability of the class s/he is teaching. Similarly, the incentive to supply effort may depend on the teacher's ability – there may be greater utility (less disutility) from effort for a high ability teacher. Thus effort may also depend on the ability fixed effect and the experience profile.

Finally, we come to the key variable, the marginal wage. The Threshold Payment increases the reward to effort for those at the appropriate point on the salary scale. Arguably, it also increases the expected reward for those below that point, but to a lower degree. We would expect that teachers would supply more effort when it was incentivised, *ceteris paribus*.

Pupil Factors

Again, the key inputs are the pupil's effort and ability. Ability depends on the history of past inputs from both school and home, as well as on the student's initial genetic capacity. [Todd and Wolpin 2001] We do not know these inputs, but like a number of other studies, we do know prior test scores and use these as a summary measure of the student's ability at the point of entry into this teaching spell. The basis for this assumption is discussed in Todd and Wolpin (2001).

We have no information on contemporaneous home inputs, and so our results on the impact of the introduction of the Threshold cannot control for this. Is this likely to be a problem? The nature of the problem depends on the direction of the relationship between parental input and teacher effort. If the two are complementary then the size of the effect will be exaggerated, but the direction will be clear. If they are substitutes, even the direction of the effect may be difficult to isolate.

[4] For example, labour supply analysis suggests that the amount of other household income will influence the supply of effort, but we do not observe this.

However, the fact that we are dealing with students aged 14 – 16 who are taught by different teachers for different subjects is important here. Indeed, some students are taught by more than one teacher for some subjects. This makes it much harder for parents to 'fine-tune' their input to the circumstances of each particular subject and teacher. To take a concrete example, parents might have to raise their input in maths and reduce it in english and science to match differing teacher eligibility. This seems unlikely; it seems more likely that home inputs are chosen on the basis of a student's overall performance, or the average teaching level that s/he faces. In this case, home inputs into education are a pupil-level effect and across-subject within-pupil variation is informative on teacher effort.

We take the prior test score as a measure of the student's ability, summarizing previous home and school inputs. Given sufficient data, we can allow the relationship between value added and initial ability to vary by different factors, including subject. Turning to the pupil's level of effort expended, we assume that this depends on her ability, on the teacher factors and on school factors. These are substituted out as above. This means for example that the full effect of "teacher ability" might include direct effects on student learning, plus indirect effects via teacher effort and pupil effort. Obviously, we can only identify, and are only interested in, the full effect.

One issue to be decided once the data are to hand is whether or not to include pupil fixed effects. Since the unit of observation is a subject-pupil-year, this is feasible with perhaps 3 observations per pupil. The advantages to doing so are that this deals with contemporaneous home inputs, and provides an additional control for pupil ability and effort, to the extent that ability is common across subjects. The disadvantage is mainly the loss of degrees of freedom, but also, the interaction with the lagged test score, and the possibility that ability is more subject-specific than general.

Analysis

First we set up the notation. Denote the test score y for pupil i at date t in subject s having been taught by teacher e as y_{iest}. The test score gain, that is, value-added, for pupil i is therefore $\Delta y_{iest} \equiv y_{iest} - y_{iest-1}$. We are interested in the quantiles of the distribution of test score gains teacher-by-teacher, and also the mean test score gain for teacher $e = E$:

$$\overline{\Delta y}_{Est} = \sum_{e=E} \Delta y_{iest}$$

We study the change in a teacher's mean test score gain between the two teaching cycles; this is denoted: θ_{es} .

We will first examine the distribution of θ_{es} in a non-parametric way, against teacher factors, particularly eligibility for the performance threshold. The logic of the approach divides teachers into three groups: those beyond the Threshold level, those at or just below the Threshold level, and those way below the level. Because of the differential relevance to them of the incentive, theory suggests the biggest effect in middle group, none in the first group, some in the last. But this first pass at the data does not control for other factors, so we will next estimate parametric regression models. The unit of observation for this is a pupil-subject-year. We have described above the variables we will include for this analysis – detailed specification of the equations awaits arrival of the data.

Identification

Our identification of the effect of the introduction of the Threshold is based on difference-in-difference principles. That is, we assume common time series shocks affect both eligible and ineligible teachers, and that membership of the eligible and ineligible groups is fixed. The key assumption is therefore that other times series factors that may confound a simple before/after study do not produce differential effects between eligible and ineligible teachers. Such common shocks could include other changes to policy or practice, and differences in the ability distribution of succeeding cohorts of pupils not picked up by the relatively broad bands of Key Stage 3. Eligibility is given by a teacher's position on the salary scale and so is indeed fixed at a point in time.

Given the link between position on the salary scale and experience, we can express our identification strategy in the following way. If the Threshold has a positive effect on pupil outcomes, this is equivalent to a shifting upwards of the outcomes-experience profile at and possibly just below, spine point 9, holding all else constant. Thus, for

identification, we assume it to be the case that while many events may change outcomes for all teachers, no other factor induces such a change in the outcome-experience profile.

Selection Issues

Only a fraction of schools are participating in our study, and these may not be a random sub-set. The potential problem is that factors raising the likelihood of school participation may also raise school performance.

We can do a number of things about this. First, as we have the aggregate population data alongside our school-based data, we can quantify any degree of non-representativeness of our participating schools. We can do this in terms of the whole distribution of pupil outcomes (levels and changes), in terms of standard school league tables, and in terms of the characteristics of the teaching staff as a whole. Second, our analysis includes school fixed effects so we can control for any simple sample selection issues by conditioning on unobserved school factors. Third, to deal with more sophisticated interaction issues (for example, between participation in the survey and the effect of the Threshold), we can estimate selection (participation) equations using past performance measures and characteristics as instruments.

6. Conclusions

We believe that the data structure and statistical approach we have outlined here is the only way to produce reliable evidence on the question of whether the introduction of the Performance Threshold has had a significant impact on the achievement of pupils. As the data acquisition progresses, we will update this paper with details of the sample.

Table 1 Sample – as at 19/5/01

Tranche	Date of approach	Number approached	Selection criteria	No. Inter-viewed	Yes
1	June-August 2000	40	Subset of the schools participating in the DFEE Value Added Pilot [O'Donogue et al. 1997]. The subset were chosen for their geographic proximity to Bristol	5	3 (7.5%)
2	September	167	Remainder of the schools participating in the Value Added Pilot	9	11 (7%)
3	December 2000-January 2001	263	All the schools in a set of nine LEAs – chosen to include a balanced number of schools of each type, some urban and some rural LEAs, some wealthy and some deprived LEAs, and geographically proximate to Bristol.	9	10 (4%)
4	February – March 2001	500	All of the schools in a set of 12 LEAs, selected for the same characteristics as those in the third tranche	2	12 (2%)
5	May 2001	618	Random sample of 18 LEAs		12 (2%)
	TOTAL	1588			48

TABLE 2 Data from schools

Information	**Level**
Class lists for year 10 in 1997/8 and year 11 in 1998/9, with pupil identifiers and teacher identifiers	pupil
Class lists for year 10 in 2000/1 and year 11 in 2001/2, with pupil identifiers and teacher identifiers	pupil
Pupil test/exam scores for Key Stage 3 in 1996/7 and GCSE 1998/9, for all English, Maths and Science subjects, with pupil identifiers	pupil
Pupil test/exam scores for Key Stage 3 in 1999/00 and GCSE 2002/03, for all English, Maths and Science subjects, with pupil identifiers	pupil
Supplementary information for each pupil: date of birth, gender, postcode. With pupil identifier	pupil
Teachers characteristics at 1 September 1999: age, gender, salary, experience, spine point, whether applied for PT. With teacher identifier.	teacher
Information about school policy: exam boards used, streaming/setting policy, pre-existing performance management system	school

Table 3 Reasons given for not wishing to participate, by those heads sending a letter or returning the postcard

Reason given for not wishing to participate in the study	Number
Special school, approached in error	2
Do not agree with the study's methodology or research design	4
Do not wish to participate given the controversy surrounding the Threshold	4
Cannot provide the data	15
New head, or head about to leave	15
Staff changes, staff shortages, or staff illness	19
Involved in OFSTED inspection or special measures	26
Involved in too many other research projects	26
Too many things on, or staff have no time	58
Did not express a reason, simply stated that not interested in participating, or not clear	409

Table 4: details of national datasets

Dataset	Source	Information contained in this dataset
Secondary Schools' Performance Tables	DFEE	Annual tables, showing for each school the number of pupils achieving particular grades in national exams, with some information about the characteristics of each school.
Value Added Tables	DFEE/QCA	Pupil-level information, showing for each pupil in an English secondary school their Key Stage 3 and GCSE results in each subject. Incomplete. Pupils cannot be identified, but schools can.
Annual Census of Schools (also known as Form 7)	DFEE	School level information, showing for each school information about the characteristics of pupils, class sizes, admissions, and the type of school.
Database of Teacher Records	DFEE	Information about the number, experience, training, salary and characteristics of each teacher in England. Individual teachers cannot be identified, but schools can.
Investors in People UK	Investors in People UK	List of schools awarded the Investors In People charter mark.

Table 5 School-level variables that will be constructed from national datasets

Number of pupils	Annual Census of Schools
Pupil age (in bands or average)	Annual Census of Schools
Pupil gender ratio	Annual Census of Schools
Number of pupils with statements of special educational needs	Annual Census of Schools
Number of pupils from minority ethnic groups	Annual Census of Schools
Number of pupils for whom English is an additional language	Annual Census of Schools
Number of pupils permanently excluded	Annual Census of Schools
Number of pupils eligible for free school meals	Annual Census of Schools
Number of pupils taking free school meals	Annual Census of Schools
Average pupil:teacher ratio	Annual Census of Schools
Average size of class	Annual Census of Schools
Number of classes in the school	Annual Census of Schools
Type of school	Annual Census of Schools
Status of school	Annual Census of Schools
Religious character of school	Annual Census of Schools
Number of qualified/unqualified teachers	Annual Census of Schools
Number of non-teaching staff	Annual Census of Schools
Number of teachers eligible for the Threshold	Database of Teacher Records
Average teacher age	Database of Teacher Records
Measures of average teacher experience (number with particular qualifications)	Database of Teacher Records
Average length of time since qualifying, for teachers in the school	Database of Teacher Records
Investor in People	Investors in People UK database
School selection policy	Performance Tables
Place in the school league tables (i.e. performance relative to national average)	Performance Tables

References

Burgess S and Metcalfe P 1999 Incentives in Organisations: A Selective Overview of the Literature with Application to the Public Sector. CMPO DP 99/016

Burgess S, Croxson B, Gregg P and Propper C 2001 The Intricacies of the Relationship Between Pay and Performance for Teachers: Do teachers respond to Performance Related Pay schemes? CMPO DP 01/35

Croxson B 2001

DFEE 1999. 1998 Value Added Pilot. *DFEE.*

DFEE 2000. Threshold Assessment Application Pack. *DfEE, March 2000*

Hay/McBer Consultants 2000. Research into Teacher Effectiveness: A Model of Teacher Effectiveness. DFEE Research Report No 216.

Hood C, Scott C, James O, Jones G, and Travers T , 1999. Regulation inside Government: Waste-Watchers, Quality Police, and Sleaze-Busters. Oxford University Press, Oxford.

Marsden D 2000. Teachers before the 'threshold'. Centre for Economic Performance, LSE Discussion Paper 454,

Marsden D and French S 1998. What a performance: performance related pay in the public services. Centre for Economic Performance, LSE, London,

Meyer RH , 1996. Value Added Indicators of School Performance. In Hanushek, E.A. and Jorgenson, D.W. (eds.)Improving America's Schools: The Role of Incentives. National Academy Press, Washington DC, 197-224.

O'Donoghue C, Thomas S, Goldstein H, and Knight T 1997. 1996 Study on Value Added for 16 - 18 year olds in England. *HMSO, London.*

Purslow, N. for the ATL 2000. Survey of attitudes on performance threshold (post application period). Association of Teachers and Lecturers (www.askatl.org.uk).

Richardson R 1999a. Performance Related Pay in Schools An Assessment of the Green Papers. LSE, A report prepared for the National Union of Teachers.

Richardson R 1999b. Performance related pay in schools: an evaluation of the Government's evidence to the School Teachers' Review Body. National Union of Teachers, London.

Richardson R 2000. Performance related pay in schools: an assessment of the NUT of the 9th STRB report, of the subsequent statement of the Secretary of State for Education, and of the recent documents pertaining to threshold applications. National Union of Teachers, London.

School Teachers' Review Body 2000. School Teachers' Review Body, Ninth Report 2000. *London: HMSO, February.*

Storey A, 2000. A leap of faith? Performance pay for teachers. Journal of Education Policy. 15, 509-523.

The Education Authorities Directory and Annual 1999 (L915). 2000 The School Government Publishing Company, Redhill, Surrey.

Thompson M, 2000. Performance management: new wine in old bottles? Professional Development Today. 3, 9-19.

Tomlinson H , 1992. Performance related pay in Education. Routledge, London.

Tomlinson H, 2000. Proposals for performance Related Pay for Teachers in English Schools. School Leadership and Management. 20.

Wragg E, Haynes G, Wragg C and Chamberlin R 2001 Performance Related Pay: The Views and Experiences of 1000 Primary and Secondary Head Teachers. University of Exeter, School of Education, Teachers' Incentives Pay Project Occasional Paper 1.