CMPO Working Paper Series No. 01/35

The Intricacies of the Relationship Between Pay and Performance for Teachers: Do teachers respond to Performance Related Pay schemes?

Simon Burgess[1]
Bronwyn Croxson[2]
Paul Gregg[3]
Carol Propper[4]

[1]CMPO, University of Bristol and CEPR
[2]CMPO, University of Bristol
[3]CMPO, University of Bristol, CEP, LSE and Council of Economic Advisors, HM Treasury
[4]CMPO, University of Bristol, CASE, LSE and CEPR

July 2001

Abstract
Performance related pay (PRP) has been introduced for teachers in the UK, as part of a drive to improve the outcome of public service and teaching in particular. For the introduced individual PRP scheme to be effective, teachers must have an effect on pupil attainment, must respond to financial incentives, must respond to performance based pay and within this to individual PRP schemes. This review examines evidence on these issues, drawing in particular from the experience of the USA. We conclude that individual PRP schemes can result in small positive gains in pupil attainment, but that the current operation of the UK scheme is more akin to a general pay rise for eligible teachers than a PRP scheme.

JEL Classification: I2, J4
Keywords: performance related pay, teachers' responses to financial incentives, public sector

Acknowledgements
We are grateful for the funding for this research provided by the Leverhulme Trust. The views expressed here, and any errors, are the responsibility of the authors.

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

CMPO is funded by the Leverhulme Trust.

1. Introduction

For the first time in recent history, the current UK government is in the process of introducing a national scheme for individual based performance related pay for teachers. Teachers' pay has been on a single scale, with nine basic increments. Under the new scheme, teachers who have reached the ninth increment are eligible to apply to pass a Performance Threshold. Their personal performance is assessed against various specific categories, including professional development and pupil attainment. If successful, they receive an annual bonus of £2000, which they will continue to receive until the end of their career, without needing to reapply. They also move on to a new, upper pay scale where they will be eligible for further performance-related increments.

This constitutes an individual performance related pay (PRP) scheme for teachers. A number of UK commentators are highly critical of the notion of PRP in education in general, and the current performance threshold scheme in particular. Teaching unions have also shown outright hostility to these proposals. Objections to the individual performance pay revolve round three overlapping concerns. First, teaching is multidimensional and aimed at much wider outcomes than exam results or test scores. Second, teaching involves team-based cooperation that is inconsistent with an individual PRP scheme. Third, teachers are professionals and do not require financial incentives to induce effort.

There are clearly questions over whether and how teachers will respond. This paper examines the evidence on this. To assess the strengths of the arguments for and against the scheme, this paper addresses four issues that are central to understanding of whether, and how, this scheme will have an impact. The first is whether teachers make a difference to children's outcomes: how important is teacher effectiveness to child educational attainment relative to other factors such as family background, peer groups, school resources etc? The second is whether teachers respond to financial incentives: even if teachers do make a difference, if they are not motivated by money then a scheme which pays them more for better pupil attainment will have no impact. The third is whether

effective individual performance related pay schemes can be operated in the public sector and how effective they have been when they have been tried. The fourth is whether PRP schemes can be used in schools and the impact of such schemes both on teachers and their pupil's outcomes.

To address these issues we survey a range of literature. In addressing the first issue, we build upon a recent comprehensive study of teachers' effectiveness (Vignoles et al. 2000) and therefore do not devote much space to this topic. In our assessment of the use of PRP schemes in the public sector we draw on international evidence, and in our survey of the impact of PRP schemes in education we draw primarily on the US experience, as this is where most schemes have been implemented. We devote attention to both teacher's perceptions of these schemes, and their impact on pupil performance, although the evidence on the latter issue is somewhat limited, although arguably the more important factor.

The layout of the paper is as follows. We begin with a brief outline of the scheme currently in use in the UK. Section 2 identifies the main findings from the large literature on the impact of teachers on child educational attainment. Section 3 reviews the limited evidence on the impact of variations in pay levels on teachers' behaviour. Section 4 briefly documents the issues surrounding the use of PRP in public sector settings. Section 5 discusses the US evidence on the impact of PRP schemes (both group and individual) in an educational setting. Finally, section 6 concludes with a discussion of the key issues relating to the use of PRP for teachers.

1. The current scheme and its background

In the late nineteenth century English school teachers were paid according to their pupils' examination results. The scheme was abandoned in 1898, partly because teachers were concentrating on teaching only more able pupils (Hood et al., 1999). During the 1980s and 1990s there has been heightened interest in education reforms, arguably in response

to concern that poor education standards were contributing to poor economic performance and the very low levels of economic activity among the least qualified (Hood et al., 1999). Since 1989 a number of commentators and organisations have expressed concern about teacher performance, and advocated the introduction of PRP for teachers (Marsden and French, 1998; Tomlinson, 1992 and 2000). In 1991, the then secretary of state for education, Kenneth Clarke, stated that he favoured the introduction of PRP for teachers (Tomlinson, 1992). The School Teachers' Review Body (STRB) was established in 1992, with a remit to "develop proposals for linking teachers' pay more closely to performance" (Marsden and French, 1998). In 1993 excellence points were introduced for teachers, nominally a form of reward for excellent performance, but few schools used them because of lack of funding and lack of an appraisal system. During the 1990s the only concrete and comprehensive PRP scheme introduced into English schools was for heads and deputies. Head teachers have had a performance related pay scheme since 1991, which has evolved over time so that pay is explicitly related to performance appraisal with at least some head teachers' targets including pupil attainment (Marsden and French, 1998; Tomlinson, 1992).

More recently, the 1997 Labour government published a Green Paper signaling its intention to introduce performance related pay for individual teachers, alongside a compulsory performance appraisal system and a number of other reforms, including a school-level performance-related rewards scheme. The current individual scheme, adopted as a result of the Green Paper, operates in the following way. Teachers have historically been paid on a single scale, with nine basic increments. Under the new scheme, teachers who have reached the ninth increment are eligible to apply to pass a Performance Threshold. Their personal performance is assessed against various specific categories, including professional development and pupil attainment. If successful, they receive an annual bonus of £2000 (which they will continue to receive until the end of their career, without needing to reapply), and move on to a new, upper pay scale where they will be eligible for further performance- related increments. The first round of this scheme began in 2000.

The schools' award is an annual competition, with schools paid awards according to their test results. Schools are required to distribute the awards to teachers as bonuses. The first awards were announced in March 2001, and 6800 schools were given up to £25,000 each (Times Educational Supplement, 23 March 2001). Performance was judged according to levels of pupil attainment (not based on value-added attainment) controlling for relative deprivation using free school meals. Awards were given in two categories: to those schools that have improved over time, and to those that excelled.

A number of UK commentators are highly critical of the notion of PRP in education in general, and the current performance threshold scheme in particular. Richardson (1999a) argues that empirical evidence about the impact of PRP in other parts of the public sector shows that it does not motivate public servants (they do not believe that it will motivate them) and that it undermines morale. He argues that the performance threshold scheme is not designed to motivate teachers but rather is designed to be a cheap way of increasing the pay of some teachers while leaving average pay unchanged. In later papers Richardson (1999b, 2000) argues that the design of the scheme is such that it will not generate better teaching performance since teachers do not believe that it is legitimate, the goals are not under the control of individual teachers and there are too many different goals. Thompson (2000), writing for a union, the Association of Teachers and Lecturers, argues that PRP *per se* is inappropriate since teachers are not motivated by extrinsic rewards but by altruism, by affiliation to their school and to their colleagues, and by personal growth.

2. How important are teachers? Issues in estimating the production of education, performance and efficiency

The importance of schools and educational resources generally in the educational attainment of children has been surprisingly controversial. The Coleman Congressional Report of 1966 (Coleman et al. 1966) concluded that schools did not make a great deal of difference to student outcomes in the US. The report looked at the wide differences in

attainment between Black and White students and came to these conclusions despite noting the large differences in school resources allocated across racial groups. The report suggested that family resources, ability and socio-economic background were the determinants of student attainment. Whilst the strength of these conclusions have been challenged since (e.g. Rutter et al. 1979), the importance of school resources to student outcomes remains ambiguous (Hanushek, 1986 , 1996).

There have been two major methodologies for looking at the impact of schools, resources and teacher inputs on student attainment. The first is simple OLS estimation of the relationship between the chosen output and the input of interest. These models look at the observed variation in outcomes such as standardised test scores, truancy rates, school completion rates etc. and assess the importance of school variation relative to aspects of family background. Reynolds et al. (1996) suggest that school quality drives around 8-12% of the variance in student achievement. Attempts to specify the aspect or characteristics of schools that lie behind measured school effectiveness have not been very precise, but patterns have emerged from meta-analyses, discussed below. The second dominant approach has been to estimate an education production function and measure how far schools fall short of the best practice as embodied in the most efficient school or group of schools. Mayston (2000a and 2000b) discusses the econometric issues involved in trying to estimate aspects of education production functions. Estimation of such frontiers can be parametric through stochastic frontier regression or they can be non-parametric Data Envelopment Analysis.

In their comprehensive and seminal paper on the existing evidence on the link between resources and pupil attainment, Vignoles et al. (2000) argue that there are three major problems with the bulk of the available literature. First, there is clear lack of any theoretical structure to guide estimation. Second, limited data availability and poor data quality lead to measurement error, a reliance on aggregated data (which in turn leads to aggregation bias) and omitted variable bias. Third, they argue that there are a host of reasons why standard estimates are plagued with endogeneity problems. These include any education funding formulae that allocates money to schools on a systematic basis that

deviates from a simple per capita formula, selection of pupils into schools via parental choice of location or allocation of resources to match student need within schools. As an example of the problem, if a school with more special needs children gets more money allocated to it, then extra resources will be negatively correlated with attainment.

There are several possible ways to address the problem of endogeneity. The first is simultaneous estimation of resource allocation and outcomes (Mayston, 1996), though identification issues can be severe. Second is the use of value-added models to net out the impact of child ability and family background on arrival at school. This can at best be described as a partial solution, as such background characteristics are still likely to be important in continued education development whilst at the school. Third, and perhaps the most promising approach, is the use of quasi-experimental approaches (instrumental variables, difference in difference estimation or random assignment experiments) to identify exogenous variation in the input factor of interest. These approaches have generally, but not universally, found positive results: extra resource inputs raise attainment. Akerheilm (1995), Angrist and Lavy (1999) and Figlio (1997) use the instrumental variable approach, and Card and Kreuger (1996) the difference in difference approach to look at the impact of changing school financial resources across racial groups in North and South Carolina, USA. Angrist and Lavy (2001) look at the impact of extra teacher training in eligible schools, compared to ineligible ones, in Jerusalem, Israel. Kreuger (1999) reports on experimental evidence on class size. The limitations of these approaches are the need for quality instruments in IV estimation and, more generally, that normally only one dimension of educational input changes at a time. This prevents the analysis of interaction effects (for example between class size and type of pupil) and also makes assessment of the relative importance of different types of input difficult.

The importance of teachers

Despite, or perhaps because of, these problems meta-analyses of the results of large numbers of studies show on balance that school resourcing does raise attainment (Dewey et al., 2000 Hedges et al., 1994). The most common positive results are for measures of

expenditure per pupil, teacher experience and teacher salary. Teacher education and other teacher characteristics are also sometimes found to be associated with attainment. Salaries make up a large proportion of school education expenditure and Hanushek et al. (1998) suggest that teachers are the most important school specific factor in influencing pupil attainment. Indeed, they suggest that 7.5% is the lower bound for total variation in attainment due to individual teachers.

But whilst teachers appear to be important in how the education system influences attainment, the findings of what it is about teachers that matters is far less clear. Teacher experience has been found to be important but the magnitudes are not large. Hanushek et al. (1998) find that experience is a strong influence on student attainment but only when comparing teachers in the first 2 years of their career with those having more than 2 years experience. Their results suggest that an experienced (2 years plus) teacher raised test scores in maths by 7 to 15% compared to a novice teacher and by 4 to 10% in reading. Kreuger (1999) finds smaller effects but the peak is somewhat later into the teacher's career. Other studies find even smaller effects or insignificant ones. Ballou and Podgursky (1999) note that teachers' salaries give high returns to seniority (even more so in the US than the UK). This evidence of a weak relationship between experience and pupil outturns and high monetary returns suggests a misalignment of rewards and teaching effectiveness. Ballou and Podgursky also argue this misalignment is not consistent with a need to keep more experienced staff in the system.

Teacher education is generally not found to be a substantial or significant influence on outturns but Angrist and Lavy (2001) do find that teacher training, once in post, raises standards in a natural experiment among schools in Jerusalem.

The design of the recently introduced UK Performance Threshold scheme was informed by research on teacher effectiveness, commissioned by the DfEE from Hay McBer consultants. The details of the methodology are not clear, but they used a combination of qualitative methods and analysis of pupil attainment scores to assess the factors affecting pupil attainment (Hay McBer, 2000). They conclude that there are three main factors

influencing pupil attainment, all under the control of individual teachers: teaching skills, professional characteristics and classroom climate. They find that neither teacher age, experience nor qualifications nor school context affect pupil attainment.

3. Does pay affect teacher behaviour?

The large literature on teacher effectiveness looks at the effect of teacher characteristics on student performance, in other words what the teacher brings to teaching, rather than what motivates teachers or how they teach. Here we focus on the research into pecuniary incentives on teachers. Teacher salary levels may influence student outcomes either through recruitment and retention of more able teachers and/or because higher wages induce greater effort.

The most developed research area on what motivates teachers is around recruitment and leaving decisions. Hanushek et al. (1999) look at the results of tests teachers take on leaving teacher training to assess whether higher paying school districts get better recruits (including both those entering the profession and those changing job). They find that those districts that pay higher wages do get teachers with higher test scores. But once a district fixed effect is included, no significant relationship is found: in other words, the association may not be a function of higher wages but of other features of the district. However, movements between schools within school districts show significant migration of teachers with higher test scores and experience to schools with higher attaining pupils and fewer minority students. While not necessarily yielding more pay, teaching in such schools might be easier or more rewarding. Black teachers were an exception to this, tending to move toward schools with more minority students. Higher paying districts did, however, suffer fewer teachers leaving teaching altogether. The results suggest that salary only had a modest impact on the quality of recruits at best. Achievement standards and racial mix within schools were far stronger influences on school choice by teachers than district level pay variation. This may suggest that teachers will prefer to teach in an easier environment or that there is greater professional satisfaction from teaching more able

students. Either way, it suggests that large compensating differentials would be needed to keep teachers (especially non-Black teachers) teaching in poor largely Black neighbourhoods.

In an attempt to assess whether pay affects who becomes teachers, Dolton and Mavromaras (1994) explore the career choices of two cohorts of graduates in 1970 and 1980. Their results suggest women are more likely to choose a career in teaching, that both cohorts were sensitive to relative pay in making career choice and this sensitivity was somewhat stronger in the 1970 cohort. Nickell and Quintini (2001) assess how teachers' (and other public sector workers') position in the pay hierarchy is related to the position in the measured ability distribution of new recruits to teaching. They examine two groups of individuals: a cohort born in 1958 (the NCDS) and a cohort born in 1970 (BSC 70). Comparing the cohorts, they conclude that declining relative pay in public services has been accompanied by a decline in the academic quality of recruits. However, as noted above, teacher effectiveness and academic ability may not be strongly related, and so this does not necessarily mean a decline over time in the quality of teachers. The view that pay levels impact on recruitment is backed up by Jacobsen (1995), who found that the starting wage affects recruitment into teaching in New York county, and that the relative wage rate affects retention rates and absentee rates among teachers in this (large) county.

Research on teacher retention by Murnane and Olsen (1989, 1990) explicitly modeled the impact of salaries and opportunity costs on the length of stay in teaching for teachers in North Carolina and Michigan. As a measure of opportunity cost, these studies used either degree subject, an ability test score, or an average salary of a graduate in the same subject who did not become a teacher. They find a positive effect of opportunity wages on teacher attrition in these states. Dolton and van der Klaauw (1999, 1996) undertake a similar study for the UK, but estimate more detailed measures of the opportunity cost, using individual wage data on teachers and data on starting wages in the non-teaching sector to explicitly estimate individual specific opportunity wages. They also distinguish between the different destinations and reasons for leaving teaching. They find that both

teacher salaries and foregone wages matter for retention. The intensity of leaving teaching for non-employment is solely influenced by teacher wages and not by wages in the outside option. On the other hand, higher opportunity wages and lower wages in the profession increase the tendency amongst teachers to switch careers. In common with other studies, they also find evidence of heterogeneity in turnover propensities, which they link to observed differences in educational background, gender, social class and ability.

In the previous section it was noted that Ballou and Podgursky (1999) argue that that experience is over-rewarded in US schools, seniority producing high pay rewards but little gain in student performance. However, this perceived misalignment of the rewards to seniority does not necessarily imply that higher teacher pay produces no impact on student outcomes other than through staff recruitment.

Teacher's salary levels are found to positively influence student outcomes by Hanushek et al. (1999), Loeb and Page (1999) and in the Dewey et al. (2000) meta-analysis. However, other studies have been unable to identify significant effects from salary levels. Loeb and Page (1999) argue that that the reason that many studies fail to find significant improvements in student outcomes and teachers' pay is because the outside option, including alternative labour market opportunities and local quality of life indicators. Hence they argue that simple cross-section results are a mixture of labour supply and demand factors. When controls for area characteristics are made, they suggest higher pay and student outcomes are positively correlated, although no causal mechanism is identified. Hanushek et al. (1999) suggest that movements in pay levels among Texas school districts were positively correlated with student value added in maths and reading. This, they suggest, was not due to improved staff retention as this was not greatly affected by the movements in pay levels. In addition, when the sample was split by school recruitment and numbers of probationary staff it was those schools with no probationary staff and no recruitment which had the greatest value added gains. This, they argue, means that improved score performance by students was not driven by the

impact of pay on staff retention and recruitment but reflects broader teacher improvements by continuing staff.

4. The operation of PRP schemes in the public sector

The scheme introduced in the UK is intended to be performance related pay, rather than a general pay increase. It is therefore useful to examine whether PRP schemes can operate in the public sector. A number of reasons have been put forward why incentive schemes will not work in the public sector. These include difficulties in measuring output and the prevalence of multi-tasking. However, some of these arguably apply to the private sector too – there are difficulties in measuring output (for example, management consultants), and multi-tasking is as important: a lot of what office workers do all day in the two sectors may actually be quite similar. One distinguishing feature is the presence of multiple principals. Dixit (2000) believes the latter is the defining characteristic of the public sector, and argues that this influences the optimal incentive structure. He uses education as an example, citing the interest of parents and children, taxpayers, potential employers, teachers and government.

There has been a good deal of empirical and theoretical work investigating the optimality of design, and the effectiveness, of incentive pay schemes. Most of this work relates to private sector schemes, and in particular to CEOs (see Prendergast, 1999, and Murphy, 1999, for surveys). Evidence relating to the public sector is scant by comparison; Burgess and Metcalfe (1999b) review the theory and evidence with regard to lessons for the public sector.

The available evidence suggests that incentive schemes have similar effects in the public sector as they do in the private sector. This can be summarised as follows: they do have an effect on behaviour (that is, workers are motivated by the incentives), the schemes are not always well designed, and they can often produce sophisticated, gaming responses and unintended behaviour. Examples for the public sector (from not a very large set)

include the work of Courty and Marschke (1997, 1999, 2001) on the effects of the incentives for the offices of the JTPA scheme, and Asch (1990) on US navy recuiters. These show both that incentive schemes do have an impact on behaviour and that part of this impact can be unintended by the organisation. Another example is Kahn, Silva and Ziliak (2001), who examine the impact of the introduction of an incentive scheme for Brazilian tax collectors. They show that the scheme had a very significant effect on fine collections per inspection – around 75% higher than they would otherwise have been.

Evidence on related issues includes the finding that public sector workers are motivated by more than just their own income (Heckman, Taber and Smith, 1996). Differences in the pattern of existence of incentive schemes between the public and private sectors are not easy to interpret but may indicate that there are inefficiently few schemes in the public sector (Burgess and Metcalfe, 1999a).

5. What is the impact of PRP in education?

Most schemes to date have been introduced in the USA, where there is considerably less central and more local control over school organisation. The literature reviewing these schemes provides some guidance as to the likely effect of PRP schemes in teaching.

The US schemes

The adoption of PRP in US schools has been through several cycles this century. It waxed after world war one, when 48% of schools had a scheme, and subsequently waned so that by 1953 there were schemes in only 4% of schools (Murnane and Cohen, 1986). Following Sputnik, there was a resurgence of interest leading 10% of schools to introduce a scheme in the 1960s, and during the 1970s interest waned again so that by 1978 fewer than 4% had schemes. A survey conducted in 1978 by the Education Research Service attributed the decline to lack of funding and to problems in implementing schemes,

including unfair evaluations, to opposition from teachers and unions, and to dissension and jealousy among teachers resulting from the schemes (Moore Johnson, 1984).

Interest in performance-based reforms rose again in the 1980s, as part of President Reagan's more general education reforms, again stimulated partly by fear that the US's relative economic performance was falling (Odden and Kelley, 1997). Concern with outcomes and performance during the 1980s also reflected concern that, although resources devoted to education had increased, pupil test scores were not improving (Ladd, 1996).

Most of the schemes introduced in the 1980s were based on the performance of individual teachers. Commentators argue that the schemes were short-lived because of problems inherent in individual PRP as well as flaws in the way schemes were implemented. One commentary attributes the demise of a South Carolina scheme, operational between 1986 and 1991, to teachers' belief that evaluations were biased, and to the inclusion of pupil attainment as a measure of performance which was both outside teacher control and allegedly open to cheating (it was alleged that teachers misrepresented results) (Clees and Nabors, 1992). Others argue that, in general, the schemes were poorly implemented in that they did not clearly define objectives, had inadequate assessment procedures and lacked credibility because they were not fully funded (Odden and Kelley, 1997; Moore Johnson, 1984). Schemes based on individual teacher performance were also criticised by these authors for engendering competition between teachers and for failing to recognise that student attainment in any one class depends on their experience in all classes.

By 1990, there were few PRP schemes. A National Centre for Education Services survey found that only 2% of teachers received individual PRP, 3% received group awards, and 16% received some type of "career ladder" award (Jacobsen, 1992).

The problem of linking measures of attainment and rewards to individual teachers is argued to have stimulated the rise, during the 1990s, of group schemes based on school-

level performance (Clotfelter and Ladd, 1986; Moore Johnson 1984). There are currently a number of such school-level schemes operating in the US. Their design varies between states and, sometimes, between the school districts within states. In some states, school schemes pay bonuses to all teaching staff in a school, in others awards go to school improvement schemes (Kelley, 1999). States in which schemes can (or have to be) translated into teacher bonuses include Colorado, Florida, Kentucky, New York, North Carolina and Texas.

The Kentucky scheme has been evaluated in a number of different studies. It was introduced in 1990, and operates on a two-year cycle (Heneman and Milanowski, 1999; Kelley, 1998). The performance of schools is judged by student performance in particular tests (Kelley and Protisk, 1997). Schools are placed in one of five different categories. Those not meeting targets are placed in low categories, and have their autonomy restricted. Top schools are paid a lump sum, which staff within the school then decide how to allocate. In the first cycle, staff in 98% of schools voted to use at least some of the award for personal bonuses. In the second cycle staff in 90% of schools voted to take bonus for "personal use". Bonuses range between $1300 - $2600 per teacher.

Charlotte-Mecklenburg is a district in the state of North Carolina. Its scheme has been operational since 1992 and has also been evaluated. Initial baselines were set for each school in 1991, and in each year each school has to improve according to targets set against the baseline, which include pupil attainment and measures of behaviour such as the number of drop outs. Rewards are assessed at school level and given to schools, which have to distribute them to teachers in fixed amounts (Heneman, 1999).

The school district of Dallas (Texas) introduced a scheme which was operational from 1993 (Ladd, 2000). Awards are calculated from the difference between the current years and past years test scores, adjusted for socio-economic variables. The method for aggregating pupil scores to a school score has been described as “incomprehensible to most participants in the process and to most outside observers… school officials neither understand the process nor have any idea what sorts of gains would have been required

for them to achieve a high ranking" (Clotfelter and Ladd, 1996). Other variables such as drop out and attendance rates are also taken into account when calculating awards. Bonuses are paid to staff in fixed amounts: teachers and principals each receive $1000 and non-professional staff receive $500. In addition, a contribution of $2000 is made to the school activity fund, and award-winning schools are invited to a gala lunch.

In other states the award goes solely to school as an improvement fund and cannot be paid as salary bonuses to teachers. Maryland operates one of these schemes, and gives awards to schools as school improvement funds (Kelley, 1999). South Carolina has also operated a school-scheme since 1984. This scheme ranks schools according to test results and attendance, judged by year-on-year gains. Schools are judged within one of four groups, defined using socio economic indicators. Within each group the top 25% get an award. In addition, schools in the bottom 75% of each group can get an award if they exceed expectations, based on the distribution of expected results for schools with their characteristics. Awards are paid in proportion to pupil numbers, typically $15,000-$20,000. In addition, award-winning schools get a waiver from state administrative restrictions (Clotfelter and Ladd, 1996).

Mississippi operates a group scheme at the level of school districts. Districts are judged according to district-wide performance on some process variables and in terms of test results, number of graduates, and number of college enrolments. High performing districts are rewarded by being made exempt from some regulations (Elmore, Abelman and Fuhrman, 1996).

The evidence on the effect of US school level schemes

A number of empirical studies have been conducted in recent years, all focusing on school-level schemes. We begin with a review of the qualitative studies that have addressed teachers' attitudes to financial bonuses in interview or questionnaire studies.

Teachers' beliefs as to the impact of financial rewards

Most studies find that teachers believe financial rewards to be appropriate and that they believe them to have a positive impact on motivation. The findings of different studies differ in the relative weight respondents give to financial (i.e. "extrinsic") rewards as opposed to peer esteem, satisfaction and other "intrinsic" rewards, and in the mechanism by which financial rewards are believed to affect behaviour.

Studies conducted by Kelley in Kentucky and in the Charlotte-Mecklenberg district of North Carolina found that teachers believed that their primary motivation came through intrinsic rewards, such as seeing students achieve and public recognition (Kelley and Protisk, 1997; Kelley, 1998 and 1999). In Kentucky, only 20% of teachers stated that the possibility of a salary bonus was a primary motivator to change teaching practice. Most teachers in the study reported in Kelley (1999) placed a higher value on intrinsic rewards, including the satisfaction of seeing improvements in pupils and the opportunity to work collaboratively. However, this study also found that most teachers believed bonuses were desirable, and when offered a choice preferred to receive a reward in form of a bonus rather than have it given to the school improvement fund. Kelley concludes that financial awards affect behaviour, but indirectly since teachers strive for the public recognition associated with the award, rather than the financial reward itself. It is possible that the motivational effects of the Kentucky scheme were undermined by memories of an earlier scheme, operating during the 1980s, which did not pay out promised bonuses. Teachers in the study reported in Kelley and Protisk (1997) remembered this, and it led some to believe that, under the new scheme, any increase in their effort would not in fact be rewarded.

A characteristic of the Charlotte-Mecklenberg scheme, believed to motivate teachers indirectly, was that teachers in interviews viewed schemes as giving them clear performance goals, in other words, clear signals of what is required (Kelley, 1999). Drawing on the psychology literature, some commentators argue that PRP schemes may motivate teachers by giving them clear information about goals and what is expected of

them, rather than motivating them directly with the promise of rewards or threat of sanctions (Kelley, 1998 and 1999). Teachers in Kentucky studies viewed fear of sanctions as a more powerful motivation for changing teaching practice than desire for rewards. The sanctions in the Kentucky scheme were for the school to be placed in a poor performing category, which teachers feared believing it would bring a negative label to the school or to themselves, a loss of professional autonomy since schools in that group were placed under external "experts", and a loss of job security (Kelley, 1998). Another study conducted in Kentucky similarly found that subjects were more conscious of fear of sanctions than they were of rewards (Elmore, Abelman and Fuhrman, 1996).

But there is clear evidence that, in addition to intrinsic rewards and fear of sanctions, teachers also believe financial rewards are important. A study conducted in Charlotte-Mecklenberg found that teachers viewed distributing school awards as bonuses to individual teachers as appropriate, although they believed that the current level of bonuses was too low (Heneman, 1999). Subjects in this study stated that they were not motivated directly by the financial rewards, but by the implied recognition - the "thank you" - implied by the bonus. These subjects stated that they were motivated primarily by helping pupils to learn. Heneman and Milanowski (1999), in another study conducted in Kentucky and Charlotte-Mecklenberg, found that when faced with 17 alternative motivators, teachers perceived bonuses to be among the top three (in Charlotte-Mecklenberg) or five (in Kentucky). They found that teachers did not believe that changing the bonus would alter their motivation, and also found, in Kentucky, lack of support for the existing scheme, possibly reflecting its poor administration. An interview-based study conducted in Colorado before the introduction of a scheme found that teachers supported the introduction of monetary incentives, and believed a PRP scheme would increase their effort (Carter and Roberto, 1992).

The studies reported above investigate teachers' beliefs about the factors affecting effort. There are also a small number of studies which attempt to look at the impact of school-level schemes on behaviour or outcomes. Two studies include analysis of what teachers believe the impact will be on pupil attainment: teachers in the *ex ante* Colorado study

believed that a PRP scheme would have a positive effect on pupil attainment (Carter and Roberto, 1992); and teachers in Kentucky, interviewed after the introduction of a scheme, believed that the scheme had led to an improvement in student writing (Elmore, Abelman and Fuhrman, 1996). A study of the factors associated with successful performance in the Kentucky scheme found successful schools changed the curriculum and improved the skills of teaching staff so that both were aligned with the goals of the assessment process, and incorporated test taking strategies into the curriculum (i.e. taught to the test) (Kelley, 1998). These changes may have resulted from the incentives given by the group scheme. Respondents in some studies certainly believed that the group scheme had increased collaboration between teachers (Kelley, 1999). Kelley's studies also found that, in Kentucky in response to the scheme, schools rewarded students for good performance or good effort. One of her studies, based on case studies in six schools, found that two high schools rewarded students with field trips and parties, and that one elementary school used part of the award money to give a grant to the local high school (Kelley and Protisk, 1997). In another study, 87% of teachers in Kentucky had created an incentive system to encourage student performance, with rewards including food before exams, extra time on trips or extra recess time, extra credit, and exemption from final exams (Kelley, 1999).

The studies found some evidence that teachers believed negative consequences had followed the introduction of school-level schemes, including stress and additional pressure (Kelley, 1999). The practice in Kentucky of teachers within a school voting on how to divide up the awards was believed to have generated internal conflict, with some teachers concerned that others were free-riding (Kelley and Protisk (1997). Some of this conflict resulted from teachers who were in the grades in which students were assessed believing they faced a disproportionate burden.

An interview-based study conducted in various US school districts found that enduring schemes were not believed to have been designed to, nor had they had the effect of, altering teaching quality (Murnane and Cohen, 1986). Rather, schemes enabled districts and schools to meet other goals, for example allowing teachers with spare time to increase their income, supporting teachers, encouraging dialogue between administrators

and teachers about issues relating to quality, and building community support for additional funding since schools used evidence that performance was being taken into account to gather support for additional funding.

Evidence on the impact of teacher incentives on student performance

There is very little evidence on the impact of teacher incentives on student outcomes, in fact there are only papers by Ladd (1999), Clotfelter and Ladd (1996) and Cooper and Cohn (1997). Ladd (1999) describes the incentive programme for schools in Dallas, and uses panel data on schools to test for effects on test scores and student drop-out rates. The scheme, introduced in 1991/2, is school-based rather than individual teacher-based and provides monetary rewards to all teachers (indeed, all staff, including janitors and secretaries) in successful schools. The details of the programme reflect quite a sophisticated approach to measuring the outcome, namely test score gains. The test score for each individual student is regressed on that student's personal characteristics (race, gender, and eligibility for free school lunch) and the residuals calculated. These residuals were then compared to the residuals the previous year and estimates of gains produced, so controlling for linear pupil effects. The mean of these across all subjects and all enrolled pupils is the school's score. The scheme used multiple measures of student outcomes from a variety of different tests in an attempt to minimise problems of "teaching to the test"; student attendance and drop-out were also factored in. Given the overall school score, around 20% of schools each year win bonuses, worth around $1000 to teachers.

Ladd's study uses a panel of school-level student test score gains across six large Texas cities, over the period 1991-1995 (availability of comparable data prevents any "before/after" comparison). The output measure used is the pass rate on maths and reading tests, thus emphasizing the bottom end of the ability distribution. The panel regressions control for common time effects and for city fixed effects rather than school fixed effects. There are also a number of school characteristics, such as racial mix and

percent disadvantaged. The results are generally positive, in that pass rates appeared to increase faster in Dallas than in other cities. However, the results are somewhat complicated by the fact that a positive Dallas effect is also found for the year before the scheme was introduced. Effects differ by sub-groups, being most positive for Hispanics and whites, and insignificant for blacks. The study does not investigate how these improvements came about, but interestingly Ladd notes a substantial increase in turnover of school principals once the scheme was in place.

Cooper and Cohn (1997) estimate both OLS and frontier production functions for South Carolina. The variables of interest for our purposes are the participation by teachers in two incentive plans. One is a purely individual scheme whereby teachers who are able to demonstrate superior levels of performance in student attendance and performance, as well as self-improvement, are awarded a bonus of around $2000. The second scheme includes a collective element (a campus component) alongside an individual teacher bonus as above. Each school district participating in the scheme used a fraction of its incentive funds for this, which is allocated to schools with high student achievement.

Boozer (1999) sets out the details of the scheme and the context in some depth. The major problem - from the point of view of evaluating incentives - is that teachers are free to apply for an award or not. They choose to participate or not in September, to become eligible for an award in the following July. In fact around 16% applied, of whom 80% were successful (Boozer, 1999). Consequently, as Cooper and Cohn put it, “It is possible, even likely, that only the most productive teachers choose to apply for an award” (pp. 320-1). Therefore, any positive effect of this variable confounds both incentive effects and selection effects.

The empirical work is based on a cross-section of 541 classes in South Carolina. The dependent variables are class median test score gains in maths and reading. So pupil effects are controlled for by the use of score gains. Other explanatory variables included are teacher experience and qualifications, class size, and pupil mix. The authors are disappointed with the results with R^2s of around 10% – 15%, and few significant

variables (including, generally, teacher experience and qualifications, and class size). The proportion of students eligible for free school meals was significantly negative. The "incentive plan" variables are significant, and raise pupil attainment but, as noted above, these actually are not at all informative about the purely motivational effects of the scheme as the study has no control for teacher heterogeneity. The purely individual plan had (with the same bonus payout) a larger and more significant coefficient than the combined individual/school scheme.

The evidence from the UK

The scheme was introduced in 2000 and therefore there is, as yet, very little direct empirical evidence about the impact of performance-related incentives on UK teachers. The evidence that does exist is either about attitudes to the scheme prior to its introduction or is about the operation of the scheme in its first year. Two studies have surveyed teacher attitudes to the Performance Threshold scheme. Marsden surveyed teachers in January 2000, after the scheme had been announced but before details had been published (Marsden, 2000). Most respondents disagreed with the principle of PRP, although it is notable that a minority believed that the proposed salary increase (£2000 per year) would stimulate increased effort. Marsden included questions designed to elicit respondents' type with respect to their commitment to their organisation. He found that most respondents have a high level of commitment, and argues that employees of this type are not likely to be motivated by extrinsic rewards.

A second survey of attitudes to the performance threshold was conducted on behalf of a teacher union, the Association of Teachers and Lecturers, shortly after teachers had completed their applications in mid-2000 (Purslow, 2000). The survey found that of the 92% of respondents eligible to apply, 80% had applied. Most of those who applied did so to obtain the salary increase or increase in pension entitlement (71%), and only 16% to gain recognition for their work. (This result is contrary to the results of the US surveys reported above, where teachers stated that they wanted financial rewards for recognition rather than desiring the money.) Of those who did not apply, 33% said they did not have

time, 27% that they disagreed with the principle and 31% because of some other change in their circumstances. About a quarter of respondents had trouble completing the pupil progress section of the form, either because they could not get appropriate statistics, their pupils were not formally tested, or because they did not believe this could be quantified.

Storey (2000) undertook an analysis of a sample of the 4000 submissions made to the government during the Performance Threshold consultation process. She found that the submissions reveal widespread opposition to the Performance Threshold. Reasons included perceived problems with measurement of that which is important in high quality teaching, perceived practical problems with assessing pupil attainment (including the lack of a robust measure of value-added and the unreliability of the key stages exams), concern about the Threshold's effects on teamwork in schools, and the perceived likelihood that it would increase bureaucracy.

The most detailed study is that of Wragg et al. (2001) who surveyed 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 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 Wragg et al, 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.

Finally, Marsden and French (1998) conducted a survey of head teachers' perceptions of their own PRP scheme, in existence since 1991. They found that heads valued both monetary and non-monetary rewards. They also found that heads were more positive about the principle of PRP if they had actually received an award, and if their school had engaged in a formal performance review. Most respondents did not believe it had affected their motivation, many because they believed their work was already of an "appropriate standard".

Discussion

This paper set out to assess whether, based on the existing evidence, an individual PRP scheme for teachers is likely to have an impact on teacher behaviour and pupil attainment. For there to be an effect, teachers must have an effect on pupil attainment, must respond to financial incentives, must respond positively to performance based pay and, within this, to individual PRP schemes.

The evidence available to address these questions is in some places somewhat limited and the results of studies of teachers can be difficult to interpret due to the difficulty of controlling for confounding factors but, given this, our conclusions are as follows. The literature on the production of education suggests that on balance teachers do have an effect on the outcomes of students. Hanushek et al. (1998) suggest that teachers are the most important school specific factor in influencing pupil attainment, and suggest a lower bound for total variation in attainment explained by individual teachers to be of the order of 7.5 percent. However, exactly what it is about teachers that matters is less clear. There is a positive impact of training and experience on pupil outcomes, but it is fairly modest, and the evidence points to a great deal of teacher heterogeneity.

The literature on the impact of financial rewards suggest that pay does have an impact on teacher behaviour. This is more on recruitment and retention than on performance once in the post, though there is some evidence from the US that pay levels do impact on teacher performance on top of feeding in through recruitment and retention of quality staff. Again, the estimated impact of pay levels is not large, and it is clear that teachers are motivated by a wide range of other factors, including the effort required to teach pupils of different abilities and the achievement standards, and racial mix, of schools.

The evidence from the literature on PRP schemes in teaching seems to be the following. First, there is debate in the US about whether PRP is, in principal, desirable. Some critics use the "contracts literature" to argue that schools do not have the characteristics associated with successful use of piece rates (Moore Johnson, 1984; Jacobsen, 1992). Critics cite the multiple goals in education, the lack of consensus about the factors leading to effective teaching and how it can be measured, and the collaborative nature of improving pupil attainment as factors mitigating against the use of PRP for teachers. Others use the psychology literature, and argue that teachers are motivated by intrinsic rather than extrinsic rewards, rendering PRP ineffective (Kelley, 1998; Moore Johnson, 1984). Some dismiss the role of incentives altogether, and argue that changing standards is insufficient to raise pupil attainment, since it is also vital to facilitate professional development and changes in teaching practice (Darling-Hammond and Falk, 1997).

Second, with the exception of Blinder (1983), most advocates of PRP argue that schools should be treated differently to other types of organisation. Advocates generally argue that education has a number of features that must be taken into account when designing and implementing schemes, but they argue that its careful implementation can lead to beneficial outcomes (Hannaway, 1996; Hanushek et al. 1994; Kemmerer and Windham, 1997; Odden and Kelley (1997). Hanushek and Jorgenson (1996) argue that the introduction of outcome-based performance incentives into US schools is vital as a way of improving performance. They argue that substantive increases in the resources devoted to education have not resulted in improved education standards. Some advocates argue that, although teachers are motivated by both extrinsic and intrinsic rewards, the former

can be used to enhance rather than necessarily undermine total motivation (Odden and Kelley, 1997).

Third, advocates and critics identify a number of issues needing to be considered when implementing PRP in schools. There are particular problems in the US education incentive structure, resulting from the nature of tests and the role they play. They argue that students are not motivated to meet attainment targets, since these do not influence employers or college admissions (Bishop, 1996; Hanushek et al. 1994). Gaming or cheating are frequently referred to, although there is little hard evidence of this type of behaviour. The Dallas school district publicised two cases of cheating, involving schools falsified exam scores (Clotfelter and Ladd, 1996) and Cless and Nabors (1992) allege that it was a factor leading to the demise of the South Carolina scheme. Other possible responses include “teaching to the test”, coaching, and changing the intake of students or focusing on high achievers to improve results (Kovetz, 1996; Ladd 1996). Several commentators discuss the importance of ensuring that incentive schemes are developed locally, with the participation of local teachers, as a means of improving the ex-post motivating effects (Carter and Roberto, 1992; Kemmerer 1997). There is widespread recognition that value added measures of pupil attainment should be used rather measures based on levels (Meyer 1996). These considerations may mean that a well designed incentive scheme will be costly to administer, as the multi-faceted nature of teaching requires a wide range on monitoring systems.

Fourth, despite these reservations, the recent evidence on the operation of schemes provides limited support for PRP for teachers. Studies of teachers views' about PRP schemes indicate that, whilst teachers place a high value on intrinsic rewards, the use of bonuses was not viewed as inappropriate. In one documented case, when offered the choice of keeping the bonuses or putting them into a school wide fund, teachers stated they would choose the former (e.g. Kelley, 1999). The evidence on the effectiveness of these schemes on (gains in) pupil attainment is more limited, but what there is suggests that such schemes have on balance a small positive effect (Ladd, 1999). The limited evidence on design (essentially group versus individual schemes) suggests that when

asked about schemes, teachers state that they prefer reward schemes that are linked to school, rather than individual, performance. However, the one study that is able to compare the impact on pupil outcomes of individual versus school level rewards for the same teachers found that the gain in pupil attainment was, in fact, larger in the individual reward component of the scheme, though the differences in magnitudes are not large, and there are problems with the scheme design (Cooper and Cohn, 1997). In either form, PRP schemes did not, despite the oppostion of some groups in teaching, appear to harm pupil attainment.

The evidence on the UK scheme (as of mid-2001) is that it has, in its first year, *ex post* operated more as a general pay increase for (almost) all teachers at the eligible point of the scale. The literature suggests that such a general pay increase may have little impact on pupil attainment, though it may help retention rates. As yet there is no evidence on whether, and how, the scheme has affected pupil performance. If the scheme continues to operate to give almost all eligible staff a pay rise, then we would not expect much impact on effort, as teachers will expect to get the bonus irrespective of whether they increase their effort. On the other hand, in future years the scheme may not operate this way, either for those who were in 2000 below the eligibility threshold, or when (the yet-to-be-finalised) once-off payments are made to those who are above the threshold. The literature we have examined here suggests that PRP schemes which give similar levels of rewards to the UK scheme, but where rewards are given selectively (with a strike rate that is lower than the present 90 percent), can improve pupil attainment. Such schemes can either be at teacher level or at school level. While the average impact of such schemes on pupil gain may not be large, it appears from the most recent evidence that it will probably be positive.

References

Akerhielm, K. (1995). Does Class Size Matter?, The Economics of Education Review, 14.

Angrist, J. and Lavy, V. (1999). Using Maimondies Rule to Estimate the Effect of Class Size on Scholastic Achievement, Quarterly Journal of Economics, 114(2)

Angrist, J. and Lavy, V. (2001). Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools, Journal of Labour Economics vol. 19 no. 2

Asch, M. (1990). Do incentives matter? The case of Navy recruiters, Industrial and Labor Relations Review, 43, pp.89-107

Ballou, D. and Podgursky, M. (1999). Seniority, Wages and Turnover Among Public School Teachers, University of Massachusetts at Amherst mimeo.

Bishop J H (1996). Signalling, Incentives and Schools Organisation in France, the Netherlands, Britain and the US. In Hanushek, E.A. and Jorgenson, D.W. (eds.)Improving America's Schools: The Role of Incentives. National Academy Press, Washington DC, 111-46.

Blinder, A (1983). The merits of merit pay. The Boston Globe. Tuesday August 23, 48.

Boozer, M. A. (1999). The Design and Evaluation of Incentive Schemes for Schools: Evidence from South Carolina's Teacher Incentive Pay Project. Mimeo, Hoover Institute.

Burgess, S and Metcalfe, P (1999a). The Use of Incentive Schemes in the Public and Private Sectors: Evidence from British Establishments. University of Bristol CMPO WP No. 99/015

Burgess S and Metcalfe P (1999b). Incentives in Organisations: A Selective Overview of the Literature with Application to the Public Sector. University of Bristol CMPO WP No. 99/016

Card, D and Kreuger, A. (1996). School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina, Journal of Economic Perspectives, vol. 10, Issue 4Carter KR and Roberto KA, 1992. A participatory process for policy development - performance pay. Education. 112, 360-369.

Clees WJ and Nabors ML (1992). Teacher incentive programs - do they make better teachers? Education. 113, 145-148.

Clotfelter, CT and Ladd, H (1996). Recognising and rewarding success in public schools. In Ladd, H.F. (ed.)Holding Schools Accountable: Performance Based Reform in Education. Brookings, Washington DC, 23-64.

Coleman, J. Campbell, E. Hobson, C. McPartland, J. Mood, A Weinfeld, F. and York, R. (1966). Equality of Opportunity, Washington DC US Government Printing Office

Cooper, S. T. and Cohn, E. (1997). Estimation of a Frontier Production Function for the South Carolina Educational Process. Economics of Education Review vol. 16 pp. 313 – 327.

Courty, P. and Marschke, G. (1997). Measuring government performance: lessons from a federal job-training program American Economic Review, May 1997, Vol.87, No.2, pp.383-388.

Courty, P. and Marschke, G. (1999). An empirical investigation of gaming responses to performance incentives. Mimeo, Economics Department, London Business School

Courty, P. and Marschke, G. (2001). The Job Training Partnership Act Incentive System In J J Heckman (ed) Performance Standards in a Government Bureaucracy. W.E. Upjohn Institute for Employment Research.

Darling-Hammond L and Falk B (1997). Using Standards and Assessments - To Support Student Learning. Phi Delta Kappan.190-199.

Dewey, J., Husted, T. and Kenny, L. (2000). The Ineffectiveness of School Inputs: A Product of Misspecification? Economics of Education Review 19(1)

Dixit, A. (2000) Incentives and Organizations in the Public Sector: An Interpretive Review. Mimeo, Princeton University, Department of Economics.

Dolton and Mavromaras (1994) Intergenerational Occupational Choice Comparisons: the case of Teachers Economic Journal, 104, 841-63

Doulton, P and van Der Klaauw (1996). Teacher Salaries and Teacher Retention, in (eds) W Baumol and W Becker Assessing Educational Practices. Cambridge Mass: MIT Press.

Doulton, P and van Der Klaauw (1999). The Turnover of Teacher Salaries: A Competing Risks Explanation. Review of Economics and Statistics 81, 543-552.

Elmore RF, Abelman CH, and Fuhrman SH (1996). The new accountability in state education reform: from process to performance. In Ladd, H.F. (ed) Holding Schools Accountable: Performance Based Reform in Education. Brookings, Washington DC, 23-64.

Figlio, D. (1997) Teacher Salaries and Teacher Quality, Economic Letters, 55(2)

Hannaway J (1996). Management Decentralisation and Performance Based Incentives: Theoretical Considerations for Schools. In Hanushek, E.A. and Jorgenson, D.W. (eds) Improving America's Schools: The Role of Incentives. National Academy Press, Washington DC, 97-110.

Hanushek E (1986) The Economics of Schooling: Production and Efficiency in Public Schools, Journal of Economic Literature, 24(3)

Hanushek E and et al (1994). Making Schools Work: Improving Performance and Controlling Costs. The Brookings Institution, Washington DC.

Hanushek E and Jorgenson D (1996). Improving America's Schools: The Role of Incentives. National Academy Press, Washington DC.

Hanushek, E. Kain, J. and Rivkin, S. (1998). Teachers Schools and Academic Achievement NBER Working Paper 6691

Hanushek E, Kain, J, and Rivkin, S. (1999). Do Higher Salaries Buy Better Teachers? NBER Working Paper 7082

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

Heckman, J., J.Smith and C.Taber (1996). What do bureaucrats do? The effects of performance standards and bureaucratic preferences on acceptance into the JTPA program, NBER Working Paper 5535

Hedges, L. Laine, R. and Greenwald, R. (1994). Does Money Matter? A Meta-analysis of Studies of the Effects of Differential Inputs on Student Outcomes Educational Researcher 23

Heneman HG and Milanowski AT, (1999). Teachers Attitudes about teacher bonuses under school-based performance award programs. Journal of Personnel Evaluation in Education. 12, 327-341.

Heneman HG (1999). Assessment of the motivational reactions of teachers to a school-based performance award program. Journal of Personnel Evaluation in Education. 12, 143-159.

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.

Jacobson SL (1992). Performance-related pay for teachers, The American Experience. In Tomlinson, H. (ed.) Performance-related pay in Education. Routledge, London.

Jacobson SL (1995). Monetary Incentives and the Reform of Teacher Compensation: A Persistent Organizational Dilemma. International. Journal of Educational Reform. 4, 29-35.

Kahn, C. M., Silva, E. C. D., and Ziliak, J. P. (2001). Performance-based wages in Tax Collection: The Brazilian Tax Collection Reform and its Effects. Economic Journal vol. 111 no. 1 pp. 188 – 205.

Kelley C and Protsik J (1997). Risk and reward: perspectives on the implementation of Kentucky's school-based performance award program. Educational Administration Quarterly. 33, 474-505.

Kelley C (1998). The Kentucky School-Based performance Award Program: School-Level Effects. Educational Policy. 12, 305-324.

Kelley C (1999). The motivational impact of school-based performance awards. Journal of Personnel Evaluation in Education. 12, 309-326.

Kemmerer FN (1997). Developing a participatory framework for local teacher incentives. In Kemmerer, F.N. and Windham, D.M. (eds.)Incentives analysis and individual decision making in the planning of education. UNESCO, International Institute for Educational Planning, Paris, 94-107.

Kemmerer FN and Windham DM (1997). Incentives analysis and individual decision making in the planning of education. UNESCO, International Institute for Educational Planning, Paris.

Kovetz D (1996). Using student assessments for educational accountability. In Hanushek, E.A. and Jorgenson, D.W. (eds.)Improving America's Schools: The Role of Incentives. National Academy Press, Washington DC, 171-96.

Kreuger, A. (1999) Experimental Estimates of Education Production Functions Quarterly Journal of Economics 114(2)

Ladd H , (1996). Holding Schools Accountable: Performance Based Reform in Education. Brookings, Washington DC.

Ladd, H. F. (1999) 'The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes' Economics of Education Review vol. 18 pp. 1 – 16.Ladd H (2000). The

Dallas School Accountability and Incentive Programme. Economics of Education Review. 18, 1-16.

Loeb, S. and Page, M. (1999) Examining the Link Between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-Pecuniary Variation, Stanford University mimeo.

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.

Mayston, D. J. (1996) Educational Attainment and Resource Use: Mystery or Econometric Misspecification?, Education Economics 4(2)

Mayston, D. J. (2000a) The Demand for Education and the Production of Local Public Goods. Mimeo University of York

Mayston, D. J. (2000b) Performance Management and Performance Measurement in the Education Sector. Mimeo University of York

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.

Moore Johnson S, 1984. Merit pay for Teachers: A Poor Prescription for Reform. Harvard Educational Review. 54, 175-185.

Murnane RJ and Cohen D, 1986. Merit pay and the evaluation problem: why most merit pay plans fail and a few survive. Harvard Educational Review. 56, 1-17.

Murnane R, and Olsen R (1989). The Effects of Salaries and Opportunity Costs on Duration in Teaching: Evidence from Michigan. Review of Economics and Statistics 71, 347-352.

Murnane R, and Olsen R (1990). The Effects of Salaries and Opportunity Costs on Duration in Teaching: Evidence from North Carolina. Journal of Human Resources 25, 106-124.

Murphy, K. (1999), Executive Compensation, in Handbook of Labor Economics, Vol. 3, O. Ashenfelter and D. Card, eds. North-Holland, Amsterdam

Nickell, S and Quitini (2001) The Consequence of the Decline in Public Sector Pay in Britain: A little Bit of Evidence, Centre for Economic Performance, LSE mimeo,

Odden A and Kelley C , 1997. Paying Teachers for what they know and do: new and smarter compensation strategies for Teachers. Sage, London.

Prendergast, C. (1999) The provision of incentives in firms, Journal of Economic Literature, vol. 37, pp.7-63

Purslow, N. 2000. Survey of attitudes on Performance Threshold (post application period). Association of Teachers and Lecturers (www.askatl.org.uk),

Reynolds, D. Dammons, P. Stoll, L. Barber, M. and Hillman, J. (1996) School Effectiveness and School Improvement in the UK, School Effectiveness and School Improvement 10(2), 193-158.

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.

Rutter M. Maughan, B. Mortimore, P. and Ouston, J. (1979) Fifteen Thousand Hours: Secondary Schools and Their Effects on Children, London, Open Books.

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

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. Performance management and school leadership. Education Journal. September.

Vignoles, A. Levacic, R. Walker, J. Machin, S. and Reynolds, D. (2000) The Relationship between Resource Allocation and Pupil Attainment: A Review, Centre for Economic Performance, LSE Mimeo

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.