British Journal of Industrial Relations

WILEY

BJIR An International Journal of Employment Relations LSE

SPECIAL ISSUE ARTICLE OPEN ACCESS

Appraisal Process, Merit Pay and Performance: Evidence From a Longitudinal Survey of School Teachers in England and Wales

David Marsden[1] | Lisa A. Sezer[2]

[1]Department of Management and Centre for Economic Performance, London School of Economics and Political Science, London, UK | [2]School of Work, Employment, Management, and Organisation, University of Leicester, Leicester, UK

Correspondence: Lisa A. Sezer (ls480@le.ac.uk)

Received: 11 October 2023 | **Revised:** 27 November 2024 | **Accepted:** 9 December 2024

Funding: The study was supported by research grants from LSE's STICERD foundation.

Keywords: merit pay | public sector | teacher pay

ABSTRACT

This study investigates how the quality of performance appraisals influences perceptions of merit pay – whether it is viewed as motivating or divisive – and its impact on achieving performance objectives. Using longitudinal survey data collected from classroom teachers in England and Wales between 2014 and 2018, and employing an instrumental variable approach, the analysis reveals that the effectiveness of merit pay in improving employee performance is closely tied to the quality of appraisal processes. Procedural fairness emerges as a key factor in fostering both motivational and divisiveness attitudes, whereas poorly designed or disengaging appraisals tend to amplify divisiveness rather than motivation, undermining the achievement of performance objectives. These findings highlight the need to carefully weigh both the potential benefits and drawbacks of merit pay systems.

1 | Introduction

Merit pay based on appraisal[1] is widely used in public sector organizations across many countries (OECD 2005a, 2012, p. 80ff.) as it is in the private sector (e.g. see Bryson, Forth, and Stokes 2017). It is defined as a '*permanent wage/salary increase granted to an employee as a function of some (typically subjective) assessment of individual employee performance*' (Gerhart, Newman, and Mikovich 2020, p. 321), and is one of the many methods used to link employee pay to their performance. In the UK, Pay Review Bodies provide guidance to the British government on public sector pay, and both highlighted the potential to enhance employee performance by shifting away from seniority-based progression (STRB 2012; DfE 2012).

How far the design of appraisal procedures and incentives contributes to improved employee performance has been a topic widely debated in the merit pay literature. This issue has gained increasing significance due to the growing need to incentivize work performance in contexts where measurement is challenging and where acceptance and positive motivation are required to effectively guide effort and influence behaviour (DeNisi and Pritchard 2006; Levy and Williams 2004). Evidence on the effectiveness of merit pay is mixed in the private sector (Heneman and Werner 2005; Rynes, Gerhart, and Parks 2005) and predominantly negative in the public sector, particularly regarding motivation (Hasnain, Manning, and Pierskalla 2012; Marsden 2010; Wenzel, Krause, and Vogel 2019), job satisfaction (e.g. Gius 2013) and other positive workplace attitudes (e.g. Bryson, Forth, and Stokes

2017). However, it also demonstrates that merit pay can enhance motivational outcomes and, even more significantly, employee performance when supported by high-quality procedures (e.g. Landry et al. 2017; Larsson, Eriksson, and Adolfsson 2022; Ogbonnaya, Daniels, and Nielsen 2017).

Personnel economists' research on merit pay focuses on performance measurement and incentive design, and based on agency theory, conceptualizes appraisals as a form of monitoring contract fulfilment (Cappelli and Canyon 2018; Lazear and Oyer 2012; Oyer and Schaefer 2010). Personnel economics research suggests that well-designed, systematic incentive pay systems with reliable and valid performance measures can enhance employee performance by motivating current staff (e.g. Fernie and Metcalf 1999a, 1999b). However, it also cautions that issues with performance measurement can diminish the effectiveness of such incentives. Measurement problems are especially pertinent in the public sector as there are often multiple and vague goals (Disney 2011, p. 225; Propper and Wilson 2003); a professionalized workforce which makes employees more resistant to tight control (Koumenta 2011); and tight budgetary restrictions. Consequently, merit pay in the public sector can result in inefficient and undesirable moral hazard effects (Brown, Potoski, and Van Slyke 2006), while also exerting pressure not only on the more technical aspects of the appraisal process, such as performance measurement and incentive design, but also on its relational dynamics.

Appraisals primarily follow a 'contractual approach' (Marsden 2010) or 'management by objectives', where employees and line managers collaboratively define goals and metrics that are meaningful for both the individual and the employer (Shadman-Pajouh 2023). Human resource management (HRM) and organizational psychology literature showed that procedural fairness (Cropanzano, Bowen, and Gilliland 2007; Thurston and McNall 2010) and involvement in organizational objective-setting (Locke and Latham 2002; Marsden and Belfield 2006) are integral motivational functions of the performance appraisal process (Latham and Pinder 2005; Levy and Williams 2004) with motivational attitudes a key outcome of merit pay practices (Larsson, Eriksson, and Adolfsson 2022; Wenzel, Krause, and Vogel 2019).

While personnel economics tends to neglect wider organizational contexts and relational aspects, which are essential to the motivational functions of the appraisal process, the organizational psychology strand makes largely only passing reference to the role of appraisal in relation to pay and performance (DeNisi and Murphy 2017; Rynes, Gerhart, and Parks 2005). This paper bridges these two approaches in the merit pay literature by assessing how both the quality of incentive design and appraisal procedures influence motivational outcomes and employees' ability to meet appraisal objectives.

Our study focuses on teachers in state primary and secondary schools across England and Wales, representing a significant and influential professional workforce within the public sector. Despite earlier evidence suggesting that performance-related and merit pay had limited motivational effects on public sector employees,[2] and opposition from teaching unions, a mandatory merit pay scheme was gradually implemented for teachers in state schools across England and Wales.

This roll-out happened progressively from 2000 (Marsden and Belfield 2006), introducing performance-based progression on a new upper pay scale,[3] to its most recent expansion in 2014.[4] The 2014 merit pay scheme sought to both expand and tighten merit pay through two main mechanisms: Firstly, it replaced seniority-based annual progression for teachers on the main pay spine with performance-based progression based on annual appraisal. Secondly, it tightened up the promotion at the 'Threshold' from the main to the upper pay spine as the Threshold system was judged to have become ineffective or rather automatic (STRB 2012).

Previous research highlighted negative attitudes among teachers towards merit pay (Hasnain, Manning, and Pierskalla 2012, p. 16), and its potential to disrupt the interdependent and collegiate teamwork in the public sector (Koumenta 2011). Therefore, it is important to evaluate the balance of both motivational and divisiveness performance effects (Wenzel, Krause, and Vogel 2019), especially given the new system's aims, based on the School Teachers' Review Body (STRB 2012), were improving teacher motivation as well as teacher recruitment and retention.

This paper draws on a unique longitudinal survey of teachers, complemented by archival data on school management characteristics, spanning 2014–2018. Following the introduction of a compulsory pay scheme in 2014, all schools were required to adopt the same payment system, eliminating selection effects that often constrain organizational case studies of specific reward systems. This provides a rare opportunity to examine broader relational appraisal practices, overlooked in many economic studies, and to explore their connection to merit pay, a link often absent from organizational psychology research.

2 | Literature: The Appraisal Process's Role in Supporting the Motivational Aspects of Merit Pay

For many years, financial incentives were deemed unsuitable for education due to challenges in measuring performance (Murnane & Cohen, 1986), the collaborative nature of teaching (Dolton, McIntosh, and Chevalier 2003) and the predominance of a 'collegiate ethos' in the sector (Adnett 2003; Koumenta 2011). However, recent studies (e.g. Atkinson et al. 2009; Pham, Nguyen, and Sringer 2021) challenge this perspective. For example, Atkinson and colleagues (2009) found that financial incentives under the 'Threshold system'[3] in England raised pupil performance, though questions about its long-term effects persist (STRB 2012). While the evidence base is still limited (Podgursky and Springer 2007), research suggests merit pay can motivate teachers, warranting further exploration of the link between merit pay and the quality of appraisal processes, particularly since the expansion of merit pay in 2014.

The measurement of teacher performance remains a contentious issue. Objective measures, such as value-added test scores, are recognized as noisy and potentially bias-inducing (Staiger and Rockoff 2010; Murnane and Papay 2010). As a result, researchers have turned to subjective measures, which complement objective metrics by capturing additional insights (Baker, Gibbons, and Murphy 1994; Rockoff and Speroni 2010; Hussain 2012). However,

given the dual role of management as evaluator and paymaster, the effectiveness of appraisal processes in supporting merit pay requires specific procedural safeguards. Although studies on pay incentives and appraisals exist, the role of appraisal quality in merit pay effectiveness remains underexplored (Lazear and Oyer 2012; Oyer and Schaefer 2010). Concerns about fairness and adverse effects from limited outcome-based incentives further complicate their application in education (Hanushek and Rivkin 2010; Figlio and Kenny 2007; Hasnain, Manning, and Pierskalla 2012).

Research from HRM and organizational psychology highlights the importance of procedural fairness and employee perceptions in appraisal systems (Heneman and Werner 2005). High-quality appraisals not only ensure fairness but also enhance motivation through employee involvement in setting goals (Gagné and Forest 2008; Locke and Latham 2002). This is particularly relevant in teaching, where goals are often context-specific and tied to local conditions (Marsden 2010). Despite these insights, comprehensive reviews of appraisal literature (DeNisi and Murphy 2017; Rynes, Gerhart, and Parks 2005) rarely link appraisals to pay and performance, reflecting a divide between transactional and relational views of employment relationships (Cappelli and Conyon 2018).

This paper bridges these perspectives by examining the appraisal process in the context of merit pay, focusing on three motivational functions: adherence to systematic procedures (Cappelli and Canyon 2018), perceptions of procedural fairness (Cropanzano, Bowen, and Gilliland 2007) and employee involvement in shaping organizational objectives (Locke and Latham 2002; Marsden 2007). We also test how the availability of financial incentives across the pay spine affects teachers' motivational attitudes by altering the expectancy of the pay award (Lawler and Suttle 1973; Kominis and Emmanuel 2007), and the effect on subsequent pay recommendations. Finally, it is crucial to examine both positive and negative attitudes. While merit pay can motivate, it may also increase divisiveness, undermining collaboration and group cohesion (Carron and Brawley 2012; Greer 2012) – factors critical in the interdependent task of teaching (Brewer et al., 2015).

2.1 | Monitoring

As Cappelli and Conyon (2018) argue, the primary focus among economists has been how to design incentives to align the work goals of agents with those of the principal. Measurement problems create opportunities for self-interested agents to engage in gaming. Information asymmetry and goal incongruence can increase agency costs, and the risk of such moral hazard effects is especially high in the public sector. Gibbons and Henderson (2012) argue that 'tit-for-tat' cooperation in the workplace can be self-sustaining provided that both parties can punish bad faith actions by the other. But when performance measures and procedures are ambiguous, it can be hard for employees to judge whether particular adverse evaluations are the result of management bad faith or genuine errors. One implication is that management should bind itself to a set of systematic, transparent appraisal procedures and measures. This makes it easier for employees to monitor the evaluation process, and compare outcomes, and also enables senior management to guard against leniency by supervisors. In practice, this has been translated into objective-setting that focuses on goals that are SMART (specific, measurable, achievable, realistic and time-bound) (ACAS 2010: 15ff.; DfE 2024: 12; NEU n.d.).

The importance of two-way monitoring can also be found in the psychological theory of expectancy (Lawler and Suttle 1973). This postulates that the strength of an employee's motivation to perform depends on the expectation of rewards that are valued by employees, and their ability to provide the necessary effort. As merit pay increasingly relies upon objective setting and appraisal, Kominis and Emmanuel (2007) propose that, in the context of extrinsic rewards, performance evaluation should be incorporated more fully into an extended version of expectancy theory. Notably, they propose a series of linkages from the employee's valuation of the reward (its valence), through objective setting and performance measurement to the expected rewards. Poorly conducted appraisals can undermine such expectations. If appraisal measures and procedures remain undiscussed, non-transparent or do not follow formal DfE guidelines, then this will lower the expectation and valence of the reward. Furthermore, if the new merit pay system results in too many teachers clustering at the top of their pay spines, and teachers believe that performance will not be rewarded competently and fairly due to an overall slowing of progression caused by budget cuts (e.g. NEU 2017), then the negative expectancy of the reward (Lawler and Suttle 1973) would counteract any potential benefits of merit pay. Therefore, we also evaluate the intensity and availability of incentives across the pay spine, and especially at the spine maxima, as opportunities may not be equally distributed.

2.2 | Procedural Justice

Many have argued and shown that *procedural justice* is critical to effective motivational linking between merit pay and appraisal (Larsson, Eriksson, and Adolfsson 2022; Latham and Pinder 2005; Lee 2019; Levy and Williams 2004). Procedural justice refers to the perceived fairness of the procedures used to determine the appraisal rating (Cropanzano, Bowen, and Gilliland 2007; Greenberg 1990). Consistent with the monitoring function, the evaluation of system components of appraisals has been the focus in 'due process' models of procedural justice (Folger, Konovsky, and Cropanzano 1992; Taylor et al. 1995; Thurston and McNall 2010), which is concerned with the extent to which employees accept and consent over appraisal standards and information. Given the relatively high level of job discretion in teaching, teacher input into individual achievable and realistic appraisal objectives is crucial for a procedurally fair appraisal system. Finally, developmental support mechanisms in appraisals are crucial to perceptions of due process because some objectives can only be achieved with training or other organizational support.

Cropanzano and colleagues (2007) argue that if practices such as appraisals are seen to be procedurally just, then employees are more likely to accept adverse outcomes. Based on 2014 US Federal Employee Viewpoint Survey data, Lee (2019) showed that negative motivational effects were mitigated when public sector employees found the appraisal process to be fair and received constructive feedback. Unfair procedures generate resistance to merit pay and to management overall, and furthermore may

disrupt the trust, cooperation and affective ties (Cohen-Charash and Spector 2001; Podsakoff et al. 1993) required for group cohesion in interlinked teachers' tasks. The consequence will be a lesser willingness for employees to engage in 'organisational citizenship behaviour' (Shapiro and Kirkman 2001).

2.3 | Involvement and Self-Determination

Employee involvement more widely is the way through which employees 'attempt to have a say and potentially influence organizational affairs relating to issues that affect their work and the interests of managers and owners' (Wilkinson et al. 2014, p. 5). The importance of giving employees a meaningful voice with regard to performance appraisal systems is well documented (e.g. Levy and Williams 2004; Marsden 2007; 2013). Input into individual appraisal objectives is an aspect of how fair the procedures are judged to be following due process appraisal models (Folger 1987). Here, we focus on the extent of teacher participation in organizational, school objective-setting. In schools, leaders are obliged to formulate school priorities in a mandatory 'School Development Plan' (SDP), which are meant to inform individual performance objectives. Schools have a choice between 'top-down' management in which organizational objectives are 'cascaded down' to the objectives of individual employees, or a more collegial approach in which employees are consulted on and co-determine organizational objectives (Marsden and Belfield 2006).

More recent research has emphasized the role of appraisal in motivating better performance by fostering employee engagement (Gruman and Saks 2011). Meaningful involvement in the appraisal process can fulfil some of employees' psychological needs and desires for autonomy and self-determination in their work, and respect for their own work values (Gagné and Forest 2008; Landry et al. 2017). The classic reviews by Locke and Latham (2002) and by Cawley, Keeping, and Levy (1998) suggest that involving teachers in determining school objectives is likely to motivate as it allows teachers to take ownership of their work objectives, and it creates a more collegial and cohesive atmosphere. The lack of involvement, on the other hand, may foster divisiveness. Wenzel, Krause, and Vogel (2019) showed that a fair, participatory and transparent appraisal design can reduce public employees' perception that they are being controlled and increase a sense of self-determination, thus fostering employee motivation. Therefore, involvement in organizational goal-setting is more likely to create a basis for shared objectives, and linking pay to their achievement can then be experienced as reinforcing this process.

3 | Methods

In this section, we introduce our survey data, present the measures of the appraisal process, instrumental, and motivational and performance outcome variables, and describe our estimation strategy.

3.1 | Data and Sample

The main source of data is an annual questionnaire survey of teachers covering the time period over which the new scheme was introduced, in 2014, until it became fully operational in 2018. The new appraisal system is based on setting objectives at the start of the school year, and reviewing achievements at the end. Data was collected in the middle of the objective setting and appraisal cycle in each year, so teachers would have known their objectives, but not yet the result of their appraisal.

The questionnaires were developed in close consultation with the teachers' unions and other stakeholder organizations, and were distributed by each union using their membership email lists to a random sample of about 10% of their membership. Teaching remains one of the most highly unionized professions,[5] and thus the effect of any bias should be very small. Each year, unions used the same distribution sample plus a refreshment sample to allow for attrition. For the panel element, teachers were asked to compose their own confidential ID. Respondents' information on their schools enabled us to link their replies to DfE school data. Over the period, we received about 12,000 usable replies. Heavy use of email surveys by the unions meant there was a good deal of response fatigue. On the whole, we received a better response from older and more senior teachers on the higher points of the school teachers' salary scale. Therefore, the results were weighted to correct for this using the School Workforce Census 2013 data on the distribution by spine point (cf. Supporting Appendix 1, Tables S6 and S7).

It is possible that our study over-estimates the degree of opposition to merit pay because discontented teachers could have been more likely to respond. On the motivational questions, our respondents were more negative than responses to an identically worded question in a Department for Education (DfE 2017) study. However, that DfE survey was based on teachers in schools whose leaders had agreed to take part, which probably creates a bias in the opposite direction as schools in difficulty likely declined to take part. As a consistency check, our survey reproduced some of the questions on commitment and relations with management from the Workplace Employment Relations Survey (WERS) 2011 survey (cf. Supporting Appendix 1, Table S8). Overall, the mean values were slightly higher in our 2014 survey compared to the 2011 WERS, for example on average, 5.6% higher on the negative attitudes; and the proportion of neutral 'hard to says' was 8.8% points higher. This suggests that our respondents were not exceptionally negative, and that the difference might be reflective of discontent and uncertainty surrounding the new pay system compared to 2011 WERS data. We also compared the share of schools graded by Ofsted inspectors with those in our sample, and they were broadly similar (cf. Supporting Appendix 1, Table S9).

3.2 | Measures of Key Variables

Appraisal functions (*APPRAISE*): We examine three key motivational functions of the appraisal process: (i) adherence to *systematic* guidelines established by the DfE to support effective monitoring; (ii) employees' perceptions of *fairness* in its implementation; and (iii) teachers' *involvement* in shaping the SDP, which defines schools' overarching priorities and serves as a framework for individual appraisals.

The questions for *systematic appraisal* were intentionally factual, designed to assess the monitoring function of appraisal (by

employees) (e.g. Cappelli and Canyon 2018). Teachers were asked to what extent their appraisal meeting followed the formal guidelines set up for the scheme (DfE 2013, 2014a). The guidelines stipulate that objective setting and reviews should work on the basis of specific objectives, which should include reference to pupil progress, classroom practice and to objectives of the SDP, that teachers should have the opportunity to discuss their objectives, know how they will be evaluated and how these will be related to pay.

Questions on *fairness* were derived from procedural justice theory (Cropanzano, Bowen, and Gilliland 2007), and required a greater degree of evaluation by respondents of both the overall fairness and of the 'due process' of appraisals followed (Folger, Konovsky, and Cropanzano 1992). Teachers were asked whether they had the opportunity in their appraisals to influence the choice of objectives, whether they felt these related to items within their control and that they could achieve, whether they thought they were fair and whether they included a measure of organizational support, such as for professional development.[6]

Finally, teachers were asked in various ways about their *involvement* in the SDP: to what extent teachers were involved in its elaboration, how well managers worked together with teachers to develop and to influence final decisions on school priorities. To avoid any further 'halo' effects, we choose questions that were the most specifically related to performance management in schools and as objective as possible.

Intensity of incentives *(INCENT)*: Variation in the intensity of merit pay incentives depends on teachers' position on their respective pay spines. Performance-based progression is organized along two pay spines, one for main grade teachers (points M1 to M6), and an upper spine for senior teachers (points U1 to U3), beyond which teachers can be promoted into the leadership grades. The lower points on each spine offer scope for progression up to the highest spine point. However, the two maximum points differ: those at the top of the main spine, M6, may progress through the Threshold if they choose, and if supported by their schools; those at the top of the upper spine, U3, can move into the Leadership group, but the number of vacancies is limited. To progress to the upper spine, teachers have to pass a 'Threshold Assessment', based on a dossier of work and an assessment of their teaching performance. If unsuccessful, teachers remain at the top of the main spine. As a result, many teachers cluster at the top of either spine, with limited opportunities for progression. The other aspect of incentive effects was the potential slowing of the overall rate of progression be of budget cuts. This was also raised in our discussions with the unions (NEU 2017).

Management quality (MGT_QUAL): Information on the quality of school leadership was obtained from the Ofsted (the Office for Standards in Education, Children's Services and Skills) national school inspectorate reports. Usually, these are undertaken on a rotating 3-yearly basis; less frequently for highly rated schools, and more frequently for those judged to have problems. Ofsted inspections follow a standard format, usually involving detailed documentation on student progress, as well as 2-day on-site visits involving interviews with parents, governors and staff, as well as some observation of lessons. The evaluation takes the form of a short report under specific headings and the award of four summary grades: 'Outstanding', 'Good', 'Requires Improvement' and 'Inadequate'. Leadership is evaluated primarily for its contribution to the school achieving its educational goals. This includes an evaluation of how robust management or governance procedures are, including assessments for performance management, and for developing teamwork and effective communication across leadership levels (Ofsted 2015). Although Ofsted does not explicitly focus on appraisals, schools with strong internal leadership procedures are likely also to have strong appraisal procedures. Several studies validated Ofsted leadership ratings and reports as a measure of general leadership and management quality (e.g. Besley, Hussain, and Machin 2006; Hussain 2012). Therefore, this study used Ofsted leadership grades as instruments for all three appraisal function measures, in addition to the below.

Management relationships (MGT_RELS): The Department for Education collects systematic information on several other factors that are relevant to school performance, notably financial, pupil progress and the school workforce. We consider a number of these for the instrumental variable (IV) analysis of Equation (1), in addition to the Ofsted leadership ratings. Rates of employee absenteeism have been shown to increase with unfair treatment (de Boer et al. 2002), and stable employment, measured by the proportion of teachers aged 50 and over, is correlated with employee commitment and fairness perceptions (e.g. Pignata et al. 2016). Finally, in smaller primary schools, shorter hierarchies and frequent contact with senior leaders provides opportunities for direct voice, and thus serves as an instrument for the 'involvement' variable.

Motivation (MP_MOTIV): Our measure of merit pay motivation draws heavily on expectancy theory (Lawler and Suttle, 1973), which proposes that employees will respond to incentives if they value them (valence), if they have scope within their jobs to increase performance (instrumentality), and if they believe that management will measure and reward it competently and fairly (expectancy). We enquire first about the valence of the proposed rewards by asking whether employees agree with the link between pay and performance, and whether they think it establishes stronger recognition of good performance, and whether it was seen to lead to a fairer allocation of pay. Whether employees perceive required job-related actions as 'instrumental' to achieving valued outcomes is a key link in expectancy theory. Thus, we ask whether teachers believe that merit pay provides an incentive to undertake the required actions: to improve/sustain their quality of teaching, and an incentive to remain a teacher. Because objective-setting and appraisal play a key role, we also asked whether merit pay gave teachers a greater incentive to focus on the objectives of their performance review. Finally, we asked about whether objective setting should include an element of pupil progress. This was one of the most controversial issues for teachers because pupil progress can vary from year to year for reasons outside individual teachers' control (Staiger and Rockoff 2010), and it was also feared that school leaders might impose objectives for pupil progress it in a formulaic way.

Divisiveness (MP_DIV): Our measure of divisiveness is best considered in relation to its opposite, group cohesion. Research identifies a positive contribution of group cohesion to group performance (see meta-analyses by Beal et al. 2003; Chiocchio and Essiembre 2009). Group cohesion has been defined as relating to a

tendency for a group to 'stick together and remain united' (Carron and Brawley 2012), comprising a willingness to work together towards a common task or goal, and a shared commitment to the group (Greer 2012). Some also emphasize the importance of group structure, notably vis-à-vis line-managers (Evans and Dion 2012). The survey instrument used by Podsakoff et al. (1993: 7), for example, includes questions of trust, cooperation and affective ties. Factors which disrupt such cohesion can contribute to perceptions of divisiveness, and by implication, weaken group performance.

We focus on employee perceptions, both before and after implementation, of how the new merit pay scheme interacts with team cohesion, including relations with those responsible for leading them. Questions on favouritism and jealousies or resentment between merit pay recipients and non-recipients address cohesiveness directly. Another question addresses perceptions of conflict with the inter-dependence of teachers' performance. The question on standards relates to a belief among some teachers that merit pay conflicts with more solidaristic notions of pay for grade and pay for competence. The remaining two address broader issues of trust in how leaders will operate the scheme: does the school have the financial resources to reward quality teaching adequately; and whether the latent objective of the scheme is to increase the quantity rather than the quality of work, as might occur with a formulaic focus on test results.

Employee Appraisal Performance *(PERF)*: The Recommendation on Pay, according to DfE regulations (DfE 2014b: 19.2.b), should be made at the conclusion of the appraisal process, reflecting whether or not a teacher is deemed to have met objectives. Normally, reviews take place at the end of the school year. The scheme includes a number of checks on possible management leniency. Once approved by the school leadership and ratified by governors, usually in the autumn, a positive recommendation would lead to pay progression.[7] Ofsted school inspectors are also expected to challenge schools that offer rates of progression that are not reflected in overall school performance.

It appears that those at the top of their pay spine may receive an informal recommendation as officials of both school leader associations mentioned that leaders could sustain morale by offering a hypothetical recommendation: 'if you were eligible for progression, I would recommend it'. This introduces a degree of fuzziness into the measure. But the use of the pay recommendation measure for our study has several advantages: the format was common for all schools; it was likely to be reviewed by others than the immediate appraiser; it had financial consequences for the school; it was a relatively objective outcome; and teachers would be familiar with the term, so could report with a minimum of ambiguity.

3.3 | Estimation Strategy

Our primary focus is to assess how motivational appraisal functions and incentives influence teachers' perceptions of merit pay as either motivating or divisive as well as their success in achieving performance goals. The first step of our model can be summarized as follows:

$$MP\_MOTIV_{it}\ (\text{or}\ MP\_DIV_{it}) = a_m + b_m APPRAISE_{it} + c_m INCENT_{it} + d_m X_{it} + u_{mit} \quad (1)$$

According to Equation (1), if employees perceive the system for setting and appraising objectives to be systematic, fair and involving (APPRAISE), and if the incentive element of the merit pay scheme is available to them (INCENT), then they will find merit pay to be motivating (or divisive) (MP_MOTIV or MP_DIV). APPRAISE gauges each of the three motivational functions of the appraisal process, being systematic, fair and involving, separately to evaluate their unique effects. We separately estimate each appraisal function to be able to analyse and compare the results using both OLS (ordinary least squares) and IV methods, addressing potential endogeneity issues. Additionally, we conduct a robustness check through a multivariate analysis using seemingly unrelated regression (Greene 2008).

The incentives provided by salary progression systems commonly involve greater intensity at different points on the pay spine, and the incentive variable takes account of such differences. Subscripts (i) and (t) refer to individual teachers, the unit of observation and the time period, respectively. The X_{it} variables include controls.[8] The sample includes very few teachers from the same schools, and standard errors were clustered by teacher to allow for responses from the same individual teachers across several years.

The most direct information on employees' perceptions and motivational effects of their appraisals comes from the survey. However, if we were to use these survey responses directly in Equation (1), there is the risk of a spurious relationship between the quality of the appraisal process and motivation because of unobserved factors that affect predictors and criteria simultaneously; a problem that is common in social sciences research involving self-ratings. Podsakoff et al. (2003), and Podsakoff, MacKenzie, and Podsakoff (2012) review a wide range of examples of common method variance (CMV), including response styles such as halo effects or acquiescence, which can lead to biased estimates. CMV is a source of potential endogeneity (Antonakis et al. 2010) with the omitted cause being a method factor (Fuller et al. 2016). Other sources of endogeneity include further omitted variables, measurement error or simultaneity (reverse causation), all of which can lead to biased estimates (Antonakis et al. 2010; Cooper et al. 2020).

To address endogeneity concerns (Antonakis et al. 2010), we employ two strategies. First, we use a system of simultaneous equations and two-stage least squares (2SLS) IV methods (Greene 2008) to examine the relationship between each appraisal function and attitudes towards merit pay. Second, we perform a series of multinomial logit estimates with lagged independent variables (cf. Equation 2) to mitigate concerns about reverse causation in performance effects. Another option to account for endogeneity in longitudinal data is via fixed effects estimators to difference out time-invariant endogenous variables (cf. Laroche 2016; Bessa, Charlwood, and Valizade 2020); however, this was not possible in the present analysis due to data limitations as most school

characteristics are stable over the timeframe between 2014 and 2018.

For the two-stage IV technique (2SLS), we first regress each function of the appraisal process (systematic, fair and involving) separately on exogenous school management characteristics using systematic data from the DfE, and then substitute the estimated values of the functions of the appraisal process in Equation (1). The first-stage estimates of the appraisal process functions (APPRAISE) were based on the following DfE variables providing the strongest instruments: Ofsted leadership grades (MGT_QUAL) were used for all three measures of the appraisal process. Additionally, data on teacher absence rates and on the proportion of teachers aged above 50, indicating workforce stability, were used as instruments for the fairness perception variable; and primary school status for the involvement variable (MGT_RELS).

An important issue in instrumental analysis is the selection of appropriate instruments for the appraisal function. In order to be useful, each of the IV described above must satisfy two essential requirements (Kennedy 2008): First, the IV must be significantly and strongly related to the predictor it represents, which is confirmed by the first-stage regression results[9] (available upon request). Second, the IV must be completely uncorrelated with the structural error term for the equation. Tests for the strength of the instruments showed robust Kleibergen−Paap *F*-statistics (Kleibergen 2007) of about 10 or more for fair and involving appraisals (Staiger and Stock 1997; Stock and Yogo 2005). For systematic appraisals, they were somewhat weaker, but this may be because the high degree of compliance reduced variation among schools; however, there is strong persistence for the other two appraisal procedural variables. Finally, a Sargan test (Sargan 1958) confirmed that over-identification was not significant: χ^2 (1) = 1.06, $p = 0.31$. The second-stage IV results and their comparison with the OLS regression are presented in Table 4.

Finally, in Equation (2), we consider whether *prior* motivation (2.1) or divisiveness (2.2) influences performance (PERF) in the *subsequent* period with a multinomial logit estimation with lagged independent variables. This also includes allowance for possible variation in the intensity of incentive at different points of the pay spine in the previous period, namely at the spine maxima (INCENT_MAX), both as a separate variable and through an interaction term with motivational attitudes.

$$\mathrm{PERF}_{it+1} = a_p + b_p \mathrm{MP\_MOTIV}_{it} + c_p \mathrm{INCENT\_MAX}_{it} + d_p \mathrm{MP\_MOTIV}_{it} * \mathrm{INCENT\_MAX}_{it} + e_p X_{it} + u_{pit} \quad (2.1)$$

$$\mathrm{PERF}_{it+1} = a_p + b_p \mathrm{MP\_DIV}_{it} + c_p \mathrm{INCENT\_MAX}_{it} + d_p \mathrm{MP\_DIV}_{it} * \mathrm{INCENT\_MAX}_{it} + e_p X_{it} + u_{pit} \quad (2.2)$$

Introducing annual lags to the motivational attitude variables is highly suitable in this study because we can correctly specify annual temporal lags. Annual lags were chosen because they are empirically relevant given appraisals run on a 1-year time period, and because they restrict the analysis to enable a larger annual sample size. Teachers who meet the objectives of their appraisal reviews should receive a recommendation on pay at the end of the school year *t*, which if validated by the leadership and governing body should lead to pay progression. We can, therefore, explore whether teachers who said they found merit pay motivating at the start of the cycle (*t*−1) went on to achieve their objectives a year later (*t*). We use a multinomial logit regression because the measure of performance comprises four distinct categories, including a positive recommendation for meeting objectives, a negative one for not meeting them, and two other categories about which we are agnostic ('not sure' and 'no recommendation'). The largest of these, 'no recommendation', was used as the comparator or reference case, and it was also the category applied to those who were at the top of their pay spines and for whom appraisers were uncertain how to proceed.

4 | Results

Insight into the working of merit pay incentives can be inferred from teachers' position on the pay spine and their progression. As part of the new 2014 system, the DfE resolved to set only the minimum and maximum points on the main and upper spines, leaving schools flexible to set pay in between. However, our survey results suggest that the overwhelming majority of schools continued to mirror the former spine points.

Table 1 shows the number of years teachers have taken to advance by one increment from the spine point they were on in 2014 or 2015 (2015 was the final year for which a 3-year interval could be calculated). Most teachers on the main spine points advanced by at least one spine point within 3 years, but several took 2 or 3 years. There was also considerable bunching at the top of both pay spines. A significant share of teachers (30%) remained at the top of the main spine (M6) for 3 years and possibly longer, and only a little more than a fifth (21%) progressed within a year to the first step on the upper spine (U1). This is in contract to the pre-2014 high success rates at the Threshold noted by the STRB (2012).

Slower progression also appears to have influenced the operation of the recommendation on pay and hence perceived incentives. In panel 2, despite receiving a positive pay recommendation, only 28% saw their pay point increase in the following year, while nearly 40% had to wait 2 or 3 years, and a third (34%) had to wait more than 3 years. Such deferred increments likely have contributed to a deterioration in motivational attitudes, especially for teachers who could have expected regular annual progression under the previous system.

The replies to questions about whether merit pay is seen as motivating or divisive are shown in Tables 2 and 3. The percentages are weighted to adjust for response differences by spine position. The overwhelming response to merit pay is negative, and this accords with an earlier review by Richardson (1999), but it is stronger than those reported by the DfE (2017). Opposition to the principle appears to have hardened over time, and the experience has generally become more negative. The biggest changes occurred shortly after the first year once teachers gained experience of its operation. In contrast, measures of divisiveness, although negative, appear to have changed less, the exception being the decline in perceived favouritism. This is in contrast with an earlier study of the first round of merit pay for school teachers introduced

TABLE 1 | Spine position and rate of pay progression.

Panel 1: 2014	Percentage (%) of teachers by spine position in 2014 advancing by 1 spine point in 1–3 years					
Initial spine point	**No progression within 3 years**	**After 1 year**	**After 2 years**	**After 3 years**	**Total %**	***N***
Unqualified teacher	—	—	—	—	—	2
M1/M2	4	22	47	27	100	45
M3/M4	2	60	34	4	100	47
M5	0	57	32	11	100	35
M6 (main spine maximum)	**30**	**21**	**37**	**12**	**100**	**81**
U1	16	23	33	28	100	119
U2	15	29	18	38	100	132
U3 (upper spine maximum)	88	3	5	4	100	305
L/LP (Leadership)	100	0	0	0	100	16
Total	45	19	20	16	100	782
Panel 2: 2015	**Percentage (%) of teachers by pay recommendation advancing by 1 spine point in 1–3 years**					
Pay recommendation in 2015 appraisal	**No progression within 3 years**	**After 1 year**	**After 2 years**	**After 3 years**	**Total**	***N***
None	72	8	10	10	100	131
Negative	59	9	23	9	100	22
Not sure if positive	45	30	10	15	100	20
Positive	34	28	19	19	100	80
Total	57	16	14	13	100	253

Note: Teachers in sample in years 2014 (panel 1) and 2015 (panel 2) for whom spine position up to 3 years later could be traced. Spine maxima in bold.

in 2000 that had found improvement in views on merit pay motivation over time (Marsden and Belfield 2006).

For subsequent analysis, these sets of motivating and divisiveness questions were combined into two indices using factor analysis.[10] Each resolved onto a single axis with a minimum eigenvalue of 1.2. The alpha coefficient of scale reliability was higher for the motivational measure at 0.82 but lower for that on divisiveness, at 0.63. The factor scores displayed a good deal of stability over time, suggesting consistency in teachers' replies, there being strong year-to-year correlations of around 0.8 for the second year, but dropping to below 0.5 by the fourth year.

The impact of appraisal practices on employee perceptions of motivation and divisiveness is shown in Table 4. The 2SLS IV regressions confirm the OLS estimates, both consistently showing positive associations between each of the three functions of the appraisal process and whether teachers perceive merit pay as motivating (Table 4, panel 1). Likewise, they both consistently show significant negative relationships with perceptions that merit pay is divisive (Table 4, panel 2). As is usual with the IV method, the overall fit of the regression is slightly lower, but the coefficients are stronger. A seemingly unrelated regression[11] confirmed the same results and patterns; the only exception being that the systematic appraisal variable becomes non-significant.

Two points stand out. First, although the coefficients of the involvement and fairness functions are broadly similar, procedural fairness emerges as the strongest predictor of motivation (and this might be due to the narrower measurement of involvement related to the SDP compared to the broader fairness measure). Second, fairness has a symmetric effect, with one standard deviation increment in fairness perception increasing motivational attitudes by a third (standardized z-coefficient: 0.337), and equally reducing divisiveness by a third (standardized z-coefficient: −0.339). However, the effects of appraisals that are considered involving, and especially of systematic appraisal are asymmetric: their presence does contribute to positive motivation, but their absence increases divisiveness to a larger extent, as reflected in the higher coefficients with 'divisiveness' than with 'motivation'.

Finally, it is worth reflecting on the importance of incentive availability, as signalled by the pay spine position. In all models, there is a negative motivational coefficient for these position dummy variables. In particular, the effect of being at the top of the two pay spines (M6 and U3) is consistently significant and strongly depresses motivation while enhancing divisiveness; a pattern which is not affected by the improved estimates of the effect of each appraisal function.

TABLE 2 | Teachers' views on whether merit pay is motivating.

	Disagree strongly	Disagree	Hard to say	Agree	Agree strongly	Total %	Change 14/18
							p-values
Is a good principle							
2014	30.9	25.5	16.6	24.6	2.4	100.0	
2018	37.9	24.3	19.5	16.3	2.0	100.0	****
It means that good teaching is properly rewarded							
2014	31.7	29.5	15.7	20.5	2.6	100.0	
2018	39.9	31.4	16.5	10.5	1.7	100.0	****
It will result in a fairer allocation of pay							
2014	40.5	39.2	12.7	7.1	0.5	100.0	
2018	45.2	35.3	13.7	5.1	0.7	100.0	*
It will make me take the objectives of my performance review more seriously							
2014	26.6	26.7	21.2	21.9	3.5	100.0	
2018	33.1	34.2	16.0	15.7	1.0	100.0	****
It will give me a real incentive to improve/sustain the quality of my teaching							
2014	40.5	31.9	19.9	7.0	0.8	100.0	
2018	43.5	38.1	12.1	5.3	1.1	100.0	****
It is good that individual teachers' pay should take some account of pupil performance							
2014	26.3	27.9	24.0	20.4	1.4	100.0	
2018	29.2	32.7	22.7	13.8	1.6	100.0	****
It makes it more attractive for me to remain a teacher							
2014	52.6	28.3	14.0	4.5	0.6	100.0	
2018	54.3	31.6	10.4	3.0	0.8	100.0	**

Note: Panel sample numbers: 2014: 1234; 2018: 1113. Statistical significance: significant at <1% ****; at <2% ***; at <5%, **; <10%, *; <20% + . A *p*-value of χ^2 test greater than 10% signifies that the change between the 2014 and 2018 survey responses could have come about by chance. Panel sample weighted to take account of higher response rate among more senior teachers using School Workforce Census 2013 distribution by spine point.

Turning to the main analysis (Equation 2), namely the effect of merit pay motivation/divisiveness on performance, we adopt a before and after approach to gauge the direction of causation, by comparing performance in the current year (t) with motivation/divisiveness in the previous year (t−1) (Table 5). The reported coefficients are relative risk ratios proportionate to the comparator case. The findings in the upper panel 1 show cumulative effects across all survey years, while panels 2 and 3 (cf. Supporting Appendix 2) show results with 1-year gaps. Cumulative effects results show that teachers who perceive merit pay to be motivating in year 1 (t−1) are 27% (1.27−1) more likely to have met their objectives and received a positive pay recommendation in year 2 (t); and 9% (1−0.91) less likely to have received a negative pay recommendation (relative to no recommendation at all, while keeping other variables constant). Similarly, for divisiveness (see right-hand column), teachers who found merit pay to be divisive in year 1 are 30% (1−0.7) less likely to receive a positive pay recommendation in year 2 than those for whom no recommendation was made. While the effect sizes of motivational and divisiveness evaluations on meeting appraisal objectives are similar, only roughly a fifth of the panel teachers had responded positively to the motivational questions (Table 2), the others being either uncertain or negative. In addition, only 55% of the panel respondents in 2014 reported advancing by one spine point within 1−3 years (cf. Table 1), while 34% of 2015 panel respondents with a positive pay recommendation reported not even progressing a spine point within 3 years.

Being on the scale maximum (M6 for the main, and U3 for the upper pay scale compared to lower points) in the previous year has the expected negative effect: pay recommendations are less likely (significant RRR coefficients below one), whether or not for progression, suggesting that any beneficial effects on performance depend on the opportunity for progression and incentive availability. We then introduce an interaction term between motivational attitudes and whether teachers were at the critical spine maxima of the Main (M6) or Upper scales (U3) at the start of the cycle. Teachers who said merit pay was motivating at the outset at M6 were about 89% more likely to have achieved a positive pay recommendation at the end of the cycle, which would involve movement to U1, thus passing the Threshold; divisiveness at M6 resulted in 32% reduced likelihood of a positive recommendation. Similarly, finding merit pay motivating in the previous year at the top of the upper spine point (U3) was 30% more likely to have achieved a positive pay recommendation than none. The much stronger coefficients for those on M6 who thought

TABLE 3 | Teachers' views on whether merit pay is divisive.

	Disagree strongly	Disagree	Hard to say	Agree	Agree strongly	Total %	Change 14/18
							p-values
Leaders will use performance pay to reward their favourites							
2014	2.4	8.2	19.3	32.6	37.6	100.0	
2018	4.4	11.7	25.9	30.3	27.7	100.0	****
It will cause resentment among teachers who feel they perform well but do not receive an award							
2014	2.7	1.4	4.4	39.0	52.6	100.0	
2018	2.1	1.5	6.7	42.0	47.8	100.0	*
For all that is said about improving teaching quality, the new pay system is simply a device to get more work done							
2014	2.4	5.2	22.6	34.0	35.9	100.0	
2018	1.9	6.0	24.0	35.9	32.1	100.0	—
The link is problematic because it is hard to relate the work done in schools to individual performance							
2014	0.9	2.9	7.5	33.2	55.4	100.0	
2018	0.4	2.6	7.6	36.7	52.7	100.0	—
It will have no effect on the quality of my work because it is already at the appropriate standard							
2014	2.2	4.6	19.6	36.4	37.2	100.0	
2018	1.7	5.5	13.6	44.4	34.9	100.0	****
The size of payments is too small to make me want to work harder to get them							
2014	3.6	10.9	45.4	26.1	14.1	100.0	
2018	3.5	11.7	27.0	33.5	24.4	100.0	****
Even if my performance is good enough, I doubt if my school can afford to reward me with a pay rise							
2014	2.4	6.4	20.3	32.6	38.3	100.0	
2018	2.8	7.2	15.1	26.0	49.0	100.0	****

Note: Panel sample numbers: 2014: 1233; 2018: 1110. Statistical significance for χ^2 tests: *p*-values at <1% ****; at <2% ***; at <5%, **; <10% *; <20% +; ns–. Sample as for Table 2.

performance pay motivating may reflect the position of teachers applying to undertake their Threshold assessment, which would lead to promotion to the upper or leadership pay scales.

Variation between schools in the strictness of reviews and progression is captured in the link between higher rates of progression in schools where Ofsted inspectors awarded lower leadership grades. Panels 2 and 3 of Table 5 show consistency in these findings across the survey years using pairs of successive years (cf. Supporting Appendix 2).

Overall, strong motivation is associated with good performance among the teachers eligible for progression, and the effect of motivation on progression at the threshold from the main scale is particularly strong. Yet, the downside is equally strong in that experiencing merit pay as divisive is associated with a reduced likelihood of meeting objectives. The temporal lag between motivational attitudes and achieving objectives strongly suggests that causation runs from the first to the second.

Two possible objections to inferring causality need to be addressed.[12] Some schools may seek to avoid morale problems by adopting a policy of 'no surprises' in the year-end appraisal by giving teachers an interim sense of how they are progressing. To check for this, we ran the same regression using a 2-year lag. This robustness check confirmed the scale of the coefficients, although the resulting smaller sample meant their statistical significance was much reduced. Secondly, it is possible that the teachers of high ability or drive may find merit pay motivating and achieve their objectives without one causing the other. To test for this, we ran another IV regression substituting estimated values for motivation in the previous year for the survey values shown in Table 5. This also showed a positive, albeit smaller (1.08) and statistically significant ($p = 0.01$) effect of prior motivation on subsequent achieving objectives.[13] Undoubtedly, both factors are at work, but these tests indicate that although they may moderate the size of the effects in Table 5, they do not undermine the general causal inference that prior positive motivation contributes to achieving objectives, and the reverse for perceived divisiveness.

5 | Discussion and Conclusion

Two main conclusions emerge. First, merit pay can be both motivating and divisive, and this study confirms that motivational effects depend on how well management operates the appraisal process (e.g. Larsson, Eriksson, and Adolfsson 2022; Wenzel,

TABLE 4 | Appraisal's three functions and their effect on merit pay attitudes: Motivation (panel 1) and divisiveness (panel 2).

Panel 1: Appraisal function, incentive intensity (pay spine position) and whether merit pay (MP) is motivating

	Systematic				Felt-fair				Involve			
MP Motivates	Coef.		Coef.		Coef.		Coef.		Coef.		Coef.	
	OLS		**IV**		**OLS**		**IV**		**OLS**		**IV**	
Systematic	0.168	****	0.238	—								
Felt-fair					0.337	****	0.310	****				
Involvement									0.215	****	0.321	***
Pay spine position												
M3_4	−0.009	—	−0.014	—	−0.016	—	0.003	—	0.021	—	0.041	—
M5	−0.214	+	−0.224	*	−0.177	+	−0.147	—	−0.088	—	−0.040	—
M6 (spine maximum)	−0.298	****	−0.301	****	−0.215	**	−0.193	*	−0.251	***	−0.230	**
U1	−0.188	+	−0.206	*	−0.103	—	−0.026	—	−0.098	—	−0.017	—
U2	−0.232	**	−0.255	*	−0.128	—	−0.090	—	−0.154	+	−0.120	—
U3 (spine maximum)	−0.405	****	−0.428	****	−0.306	****	−0.273	****	−0.275	****	−0.240	***
Leadership	−0.066	—	−0.103	—	−0.030	—	0.018	—	−0.000	—	0.012	—

(Continues)

Krause, and Vogel 2019). All three motivational functions of the appraisal process, being systematic, felt-fair and involving, are pertinent. However, high-quality appraisals involve more than effective monitoring, a function highlighted in agency theory (Cappelli and Conyon 2018). Procedural fairness emerges as the best predictor of both motivation and divisiveness, although its effect is of similar size to that of involvement in setting school objectives. This matters because our analyses showed that prior motivation or divisiveness affects the likelihood of achieving objectives.

Second, we find an asymmetry of the appraisal functions' effects in that poorly designed or disengaging appraisals tend to amplify divisiveness rather than motivation. Therefore, the benefits from positive motivation among a minority are offset by the missed objectives resulting from divisiveness felt by the majority, especially for teachers at the spine maxima for whom the Threshold assessment was tightened, making the overall balance sheet negative. This might be explained by the fact that teachers feel that previously more automatic progression at the Threshold is a loss, and losses generally outweigh equivalent gains in employee perceptions (Geys et al. 2023; McGraw et al. 2010).

The implication for practice is that high-quality appraisal procedures for those eligible for merit-based progression can have a positive effect and help overcome some of the inherent difficulties of performance measurement in the public sector. However, results highlight the negative effects of merit pay systems and join a line of studies that caution against the use of individual performance-related pay types in the public sector (e.g. Bryson, Forth, and Stokes 2017; Wenzel, Krause, and Vogel 2019); and especially in teaching and professions where performance is hard to measure and group cohesion is important. The negative experiences of slowed progression and declining evaluations found in our study are likely to fuel continued divisiveness with negative performance implications of merit pay. Indeed, the STRB (2023, 2024) and DfE regard changes to appraisal and pay progression system as a priority, and the STRB (2023) report concluded that 'the current model is not working and evidence of its efficacy is limited' (p. 22).

The study has a number of limitations. First, future research could try to address the data limitations of this survey study. We were not able to undertake a school level of analysis. However, by increasing responses within each school, multi-level hierarchical estimation models could be applied (cf. Shrout and Fleiss 1979). Furthermore, the nature of the variables and data in our study prevented a fixed-effects estimation or a direct test of mediator effects because both continuous and polytomous outcome variables were manipulated. While we were limited to a robust alternative, namely a series of simultaneous equations followed by a multivariate logit estimation with lagged variables, further research could build on our work through a direct test of mediator effects, for example with a conditional process model (Hayes 2012). There is a need for further longitudinal studies to evaluate carefully the long-term effects of negative merit pay evaluations in the context of a squeezed public sector with recruitment and retention difficulties. Future research could evaluate the effects of merit pay procedural quality on group and school performance, as well as effects on teacher turnover intentions.

Second, it is sometimes argued that education is a case apart from other economic activities, with a high reliance on intrinsic motivation and a high degree of job autonomy. Yet, intrinsic motivation is not confined to education, nor is it confined to public services, and the general debate about the interaction

TABLE 4 | (Continued)

Panel 1: Appraisal function, incentive intensity (pay spine position) and whether merit pay (MP) is motivating												
	Systematic				**Felt-fair**				**Involve**			
Female	0.129	**	0.123	**	0.135	****	0.133	***	0.125	***	0.105	*
Year	−0.024	+	−0.022	+	−0.016	—	−0.015	—	−0.026	*	−0.029	**
constant	47.569	+	44.110	+	31.738	—	30.536	—	52.707	*	57.803	**
Number of observations	3503		3503		3505		3189		3958		3584	
Prob > χ^2	0.000		0.000		0.000		0.000		0.000		0.000	
R^2	0.0544		0.0499		0.1423		0.1369		0.0725		0.0577	
Clusters (teachers)	1775		1775		1776		1671		1893		1791	
Panel 2: Appraisal function, incentive intensity (pay spine position) and whether merit pay (MP) is divisive												
	Systematic				**Felt-fair**				**Involve**			
MP Divisive	Coef.		Coef.		Coef.		Coef.		Coef.		Coef.	
	OLS		**IV**		**OLS**		**IV**		**OLS**		**IV**	
Systematic	−0.218	****	−0.435	**								
Felt-fair					−0.339	****	−0.384	****				
Involvement									−0.291	****	−0.369	****
Pay spine												
Position												
M3_4	−0.088	—	−0.067	—	−0.100	—	−0.193	*	−0.126	+	−0.215	**
M5	0.059	—	0.090	—	−0.001	—	−0.099	—	−0.098	—	−0.217	*
M6 (spine maximum)	0.280	****	0.292	****	0.186	**	0.093	—	0.203	**	0.117	—
U1	0.130	—	0.192	*	0.024	—	−0.088	—	0.024	—	−0.091	—
U2	0.237	***	0.312	***	0.113	—	−0.001	—	0.137	+	0.036	—
U3 (spine maximum)	0.369	****	0.444	****	0.252	****	0.173	**	0.209	****	0.132	+
Leadership	0.039	—	0.156	—	−0.016	—	−0.107	—	−0.062	—	−0.144	—
Female	−0.080	+	−0.061	—	−0.096	**	−0.079	+	−0.052	—	−0.023	—
Year	−0.025	*	−0.031	**	−0.033	***	−0.036	***	−0.020	+	−0.018	+
Constant	51.083	*	62.719	**	65.573	***	72.722	***	40.842	+	37.155	+
Number of observations	3508		3508		3510		3192		3957		3581	
Prob > χ^2	0.000		0.000		0.000		0.000		0.000		0.000	
R^2	0.0702		0.0254		0.1409		0.1368		0.1074		0.0965	
Clusters (teachers)	1777		1777		1778		1672		1893		1789	

Note: Coefficients are standardised z-coefficients for continuous variables. Spine point effects, as explanatory variables, take junior teachers as the benchmark (unqualified and on points M1 and M2). The categories M1-2 and M3-4 are merged because of small sample numbers. Weighted estimates; standard errors clustered by teacher to allow for responses from the same individual teachers across several years. Control variables included year and gender (dummy variable if 'female'). Some studies suggest that merit pay has a gender dimension with women less likely to find it acceptable and fair (Dohmen and Falk 2011); however, we find that female gender was associated with more positive motivational attitudes. Statistical significance: significant at <1% ****; at <2% ***; at <5%, **; <10% *; <20% +; ns–.

TABLE 5 | The effect of finding merit pay motivating or divisive on subsequent performance.

Panel 1: all years combined								
	MP_Motivating (*t*−1)						**MP_Divisive (*t*−1)**	
	Performance: Current pay recommendation (*t*)						**Performance: Current pay recommendation (*t*)**	
	Negative		**Not sure**		**Positive**		**Positive**	
MP motivating/divisive at *t*−1	0.911	—	0.887	—	1.272	****	0.700	****
On M6 at *t*−1	1.017	—	0.954	—	0.556	****	0.575	****
MP motivating/divisive and on M6 at *t*−1	1.181	—	2.172	****	1.887	****	0.681	*
On U3 at *t*−1	0.369	****	0.183	****	0.085	****	0.082	****
MP motivating/divisive and on U3 at *t*−1	0.915	—	1.336	—	1.301	*	0.828	—
Ofsted ldr grade 2	1.259	—	0.749	—	1.244	+	1.278	*
Ofsted ldr grade 3	0.887	—	0.529	—	1.796	****	1.938	****
Ofsted ldr grade 4	1.686	—	1.893	—	0.974	—	1.055	—
Year	0.689	****	0.195	****	0.688	****	0.675	****
Constant			.		.		.	
Number of observations					2674		2675	
Pseudo R^2					0.1419		0.1461	
Clusters (teachers)					1646		1644	

Note: Table shows relationship between whether merit pay was said to motivate (left column) or be divisive (right column) in previous survey year (*t*−1), and the appraisal performance measure: whether no (base case), a negative or positive pay recommendation was received (or not sure) in the following year (*t*), indicating whether appraisal goals were met. Standardized z-scores estimated for 'MP motivates' and 'divisive' via multinomial logit regression. Coefficients show relative risk ratios (RRR) of pay recommendation outcome compared with base case of no pay recommendation = 1. Clustered standard errors by teacher ID. Weighted regression. Statistical significance: significant at <1% ****; at <2% ***; at <5%, **; <10% *; <20% +; ns–.

between intrinsic and extrinsic motivators has been set in the context of the economy as a whole (Gagné and Deci 2005). Likewise, many occupations other than teaching involve highly qualified workers where monitoring is costly. What is shared is a common difficulty for management to monitor employee effort.

Third, being located in the public sector limited the scope for funding merit pay, and this was compounded by the austerity programme throughout the period. We argue that the deterioration between 2014 and 2018 of already low teacher merit pay motivation is a result of the operation of merit pay. It is these school variations that are picked up in our regression analyses rather than the background effect of austerity. The 2014 figures give a rough picture of teachers' views when austerity would have been linked to general pay freezes, but before merit pay gave schools the additional means to manage their staff costs by slowing pay progression. Nevertheless, the tight budgetary controls on schools throughout the period of this study mean that it may underestimate some of the potential benefits for organizations of merit pay and greater pay flexibility (e.g. Lazear 1998; Ogbonnaya, Daniels, and Nielsen 2017) especially in the private sector. On the other hand, even private organizations require that pay reforms should be cost-neutral as the potential gains from increased motivation have yet to be realized.

A final implication is that the findings accord with long-term employment as a relational exchange, in which employers and employees need to feel fairly treated as their agreement gradually evolves (Marsden 1999; Cappelli and Conyon 2018). Both parties have a stake in the continuing exchange, and so need to be able to agree on how to adapt over time. As performance is hard to measure and highly qualified employees have their own views on relevant measures, appraisal can provide a forum for implicit, integrative negotiation (Marsden 2007; Marsden and Belfield 2006) in which both parties agree on objectives and how to evaluate them. Appraisal can serve as a channel for employee voice at the individual level, which may contribute not only to improved school performance, but also help mitigate negative effects on teachers' motivation, but only if implemented in a systemic, felt-fair way that involves teachers in linked up goal-setting. The negative effects of the merit pay system could be further dampened if complemented by collective voice through teachers' unions (Della Torre 2019). Performance management more widely, and merit pay in particular, threatened the scope of traditional union bargaining activities. Earlier findings (e.g. Belfield and Marsden 2004) argued that merit pay could offer a new institutional role to unions as providers of 'procedural fairness services' to both union members and employers to provide advice on implementation and to address the multiple concerns over divisiveness, workload and budget cuts. This appears especially important in teachers' merit pay as opposition and negative evaluations have increased,

with negative implications for teacher retention during teacher shortages.

Acknowledgements

This paper has benefited from advice and support from the stakeholder organizations in school teaching in England and Wales, and notably the teachers' unions which distributed the questionnaires and helped pilot its contents. It has benefited from grants from LSE's STICERD foundation and the project extends LSE's Centre of Economic Performance research on performance pay. The views expressed in this paper are those of the authors and do not necessarily reflect those of any of the above organizations. Acknowledgements are also due for use of the 2011 Workplace Employment Relations Survey results for benchmarking purposes. Furthermore, we are grateful for comments received at the BUIRA, BJIR and ILERA conferences.

Last but not least, as corresponding author, I would like to express my gratitude to the editors of the special issue, the anonymous reviewers, to Phil Almond, Sara Connolly, Maria Koumenta, and Alice Lam for their support and feedback. I feel privileged to have worked with and learnt a tremendous amount from David. I will remember David as a truly interdisciplinary scholar, kind and honourable colleague, inspiring teacher and supportive mentor to the next generation.

Conflicts of Interest

There are no conflicts of interest to report.

Data Availability Statement

The data underlying this article cannot be shared due to the privacy of the individuals who participated in the study.

Ethics Statement

All subjects participated voluntarily with informed consent. The project received ethical approval from The London School of Economics.

Endnotes

[1] 'Performance appraisal is a discrete, formal, organizationally sanctioned event, usually not occurring more frequently than once or twice a year, which has clearly stated performance dimensions and/or criteria that are used in the [performance] evaluation process' (DeNisi and Pritchard 2006, p. 254).

[2] Early studies evaluating the first large-scale performance pay schemes in the British civil service in the late 20th century failed to show widespread positive motivational effects due to negative perceptions of how the appraisals were operated, with unsystematic, top-down objective-setting lacking employee involvement in the process, as well ineffective incentives (e.g. Dowling and Richardson 1998; Heery 1998; Marsden and Richardson 1994).

[3] The 2000 system had also introduced the 'Threshold' assessment based on teaching competence for the transition from the Main to the Upper Scale. For headteachers, performance pay elements had been introduced even earlier, in 1995.

[4] From September 2021, the Welsh Government abolished merit pay in favour of seniority-based pay progression and reintroduced national statutory pay scales.

[5] A report commissioned by the Department for Education based on 2012 survey data found that all but 3% of the responding teachers belonged to a teaching union (Ager and Pyle 2013). The TUC website provides membership information excluding the headteachers' union, ASCL, and Voice membership and thus estimates based on such data will produce a lower figure. For 2014, based on total membership as reported on the TUC website, we estimate a density figure of 92%: ATL (127.9k), NASUWT (294.2k), NUT (326.9k), UCAC (3.9k) and NAHT (28.5k) (total 777.5k), divided by total teaching FTE in the 2014 School Workforce Census (848.8k).

[6] For subsequent analysis, these sets of questions were combined into three indices using factor analysis. We conducted exploratory factor analyses (EFAs) based on the common factor model with an oblique rotation. Each resolved onto a single axis with a minimum eigenvalue of 1.0. The alpha scale reliability coefficients for the three indices on systematic, fair and involving appraisals were 0.52, 0.82 and 0.84, respectively. The relatively low value of the systematic index may reflect the relatively high degree of compliance by schools with the letter of the appraisal guidelines. The factor scores displayed a good deal of stability over time, suggesting consistency in teachers' replies.

[7] In their written-in replies, some respondents complained that their positive recommendation had been overruled by school leaders or governors.

[8] Year and gender ('female') as dummy variable was included as some studies suggested that merit pay has a gender dimension with men more likely to find it acceptable and fair (Dohmen and Falk 2011).

[9] The first-stage regression results show that Ofsted leadership grades were significant across the board; with a declining relationship with appraisal quality as one moved down the leadership grades. Teacher absence, primary school status and teaching workforce stability were also significant: as expected, teacher absence was negatively, and workforce stability as well as primary school status positively related to each appraisal function.

[10] More specifically, we conducted EFAs based on the common factor model with an oblique rotation. EFA is particularly well suited to measurement instrument construction (Fabrigar and Duane 2011, p. 35).

[11] The instrumental variables were regressed on the functions of the appraisal process (systematic, fair and involving), and those on the motivational perceptions, together with incentive/pay spine variables. Results are available from the corresponding author on request.

[12] A supplementary robustness check on the linkages between fair appraisals, finding MP motivating/divisive and achieving objectives was undertaken using IV. We regressed motivation in year 2 on fairness of appraisal in year 1, and achieving objectives in year 2 on motivation in year 1. Thus, at each step, a 1-year time gap was used between the attitude and its outcome. This overlapping method was used because it required data from only 2 continuous years, thus enabling a larger sample. The results confirm those of Tables 4 and 5. Fairness of appraisal in the first year has a positive influence on motivation in the second year, and motivation in the first year has a positive effect on achieving objectives in the second year, and conversely for divisiveness.

[13] These coefficients are not directly comparable with Table 5 because it is based on an IV probit estimation and the dependent variable takes the value of 1 for a positive recommendation and 0 for all others.

References

ACAS. 2010. *How to Manage Performance*. Advisory Booklet Ref: B21. London: HMSO.

Adnett, N. 2003. "Reforming Teachers' Pay: Incentive Payments, Collegiate Ethos and UK Policy." *Cambridge Journal of Economics* 27, no. 1: 145–157.

Ager, R., and K. Pyle. 2013. "Teacher Voice Omnibus: Understanding Union Membership and Activity." Department for Education.

Antonakis, J., S. Bendahan, P. Jacquart, and R. Lalive. 2010. "On Making Causal Claims: A Review and Recommendations." *Leadership Quarterly* 21, no. 6: 1086–1120.

Atkinson, A., S. Burgess, B. Croxson, et al. 2009. "Evaluating the Impact of Performance-Related Pay for Teachers in England." *Labour Economics* 16, no. 3: 251–261.

Baker, G., R. Gibbons, and K. J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics* 109, no. 4: 1125–1156.

Beal, D. J., R. R. Cohen, M. J. Burke, and C. L. McLendon. 2003. "Cohesion and Performance in Groups: A Meta-Analytic Clarification of Construct Relations." *Journal of Applied Psychology* 88, no. 6: 989–1004.

Belfield, R., and D. Marsden. 2004. Unions, Performance-Related Pay and Procedural Justice: the Case of Classroom Teachers. CEP Discussion Paper No. CEPDP0660. London: London School of Economics.

Besley, T., I. Hussain, and S. Machin. 2006. Do Public Sector CEOs Make a Difference? Results From a Pilot Survey of Headteachers. Centre for Economic Performance Working Paper. London: London School of Economics.

Bessa, I., A. Charlwood, and D. Valizade. 2020. "Do Unions Cause Job Dissatisfaction? Evidence From a Quasi-Experiment in the United Kingdom." *British Journal of Industrial Relations* 59, no. 2: 251–278.

Brewer, T. J., P. S. Myers, and M. Zhang. 2015. "Islands Unto Themselves: How Merit Pay Schemes May Undermine Positive Teacher Collaboration." *Critical Questions in Education* 6, no. 2: 45–54.

Brown, T., M. Potoski, and D. M. Van Slyke. 2006. "Managing Public Service Contracts: Aligning Values, Institutions and Markets." *Public Administration Review* 66: 323–331.

Bryson, A., J. Forth, and L. Stokes. 2017. "How Much Performance Pay is There in the Public Sector and What Are Its Effects?" *Human Resource Management Journal* 27, no. 4: 581–597.

Cappelli, P., and M. J. Conyon. 2018. "What Do Performance Appraisals Do?" *Industrial and Labor Relations Review* 71, no. 1: 88–116.

Carron, A. V., and L. R. Brawley. 2012. "Cohesion: Conceptual and Measurement Issues." *Small Group Research* 43, no. 6: 726–743.

Cawley, B., L. Keeping, and P. Levy. 1998. "Participation in the Performance Appraisal Process and Employee Reactions: A Meta-Analytic Review of Field Investigations." *Journal of Applied Psychology* 83, no. 4: 615–633.

Chiocchio, F., and H. Essiembre. 2009. "Cohesion and Performance: A Meta-Analytic Review of Disparities between Project Teams, Production Teams, and Service Teams." *Small Group Research* 40, no. 4: 382–420.

Cohen-Charash, Y., and P. E. Spector. 2001. "The Role of Justice in Organizations: A Meta-Analysis." *Organizational Behavior and Human Decision Processes* 86, no. 2: 278–321. https://doi.org/10.1006/obhd.2001.2958.

Cooper, B., N. Eva, F. Z. Fazlelahi, A. Newman, A. Lee, and M. Obschonka. 2020. "Addressing Common Method Variance and Endogeneity in Vocational Behavior Research: A Review of the Literature and Suggestions for Future Research." *Journal of Vocational Behavior* 1, no. 21: 103472.

Cropanzano, R., D. E. Bowen, and S. W. Gilliland. 2007. "The Management of Organizational Justice." *Academy of Management Perspectives* 21, no. 4: 34–48.

de Boer, E. M., A. B. Bakker, J. E. Syroit, and W. B. Schaufeli. 2002. "Unfairness at Work as a Predictor of Absenteeism." *Journal of Organizational Behavior* 23, no. 2: 181–197.

Della Torre, E. 2019. "Collective Voice Mechanisms, HRM Practices and Organizational Performance in Italian Manufacturing Firms." *European Management Journal* 37, no. 3: 398–410.

DeNisi, A. S., and K. R. Murphy. 2017. "Performance Appraisal and Performance Management: 100 Years of Progress?" *Journal of Applied Psychology* 102, no. 3: 421–433.

DeNisi, A. S., and R. D. Pritchard. 2006. "Performance Appraisal, Performance Management and Improving Individual Performance: A Motivational Framework." *Management and Organization Review* 2, no. 2: 253–277. https://doi.org/10.1111/j.1740-8784.2006.00042.x.

DfE. 2012. *Evidence to the STRB: the Case for Change*. London: Department for Education.

DfE. 2013. *Departmental Advice: Reviewing and Revising Your School's Approach to Teachers' Pay*. London: Department for Education.

DfE. 2014a. *Implementing Your School's Approach to Pay: Departmental Advice for Maintained Schools and Local Authorities*. London: Department for Education.

DfE. 2014b. *School Teachers' Pay and Conditions Document 2014 and Guidance on School Teachers' Pay and Conditions*. London: Department for Education.

DfE. 2017. *Evaluation of Teachers' Pay Reform: Final Report*. London: Department for Education.

DfE. 2024. *Teacher Appraisal: Guidance for Schools*. London: Department for Education.

Disney, R. 2011. "The Future of Public Sector Pay in the United Kingdom." In *Employment in the Lean Years: Policy and Prospects for the Next Decade*, edited by D. Marsden, Chapter 15, 213–233. Oxford: Oxford University Press.

Dohmen, T., and A. Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences and Gender." *American Economic Review* 101, no. 2: 556–590.

Dolton, P., S. McIntosh, and A. Chevalier. 2003. *Teacher Pay and Performance*. London: Bedford Way Papers.

Evans, C. R., and K. Dion. 2012. "Group Cohesion and Performance: A Meta-Analysis." *Small Group Research* 43, no. 6: 690–701.

Fabrigar, L. R., and T. W. Duane. 2011. *Exploratory Factor Analysis*. Oxford University Press.

Fernie, S., and D. Metcalf. 1999a. "It's Not What You Pay It's the Way That You Pay It and That's What Gets Results: Jockey's Pay and Performance." *Labour* 13, no. 2: 385–411.

Fernie, S., and D. Metcalf. 1999b. "(Not) Hanging on the Telephone: Payment Systems in the New Sweatshops." *Advances in Industrial and Labor Relations* 9: 23–67.

Figlio, D. N., and L. W. Kenny. 2007. "Individual Teacher Incentives and Student Performance." *Journal of Public Economics* 91, no. 5–6: 901–914.

Folger, R. 1987. "'Distributive and Procedural Justice in the Workplace'." *Social Justice Research* 1, no. 2: 143–159.

Folger, R., M. A. Konovsky, and R. Cropanzano. 1992. "A Due Process Metaphor for Performance Appraisal." *Research in Organizational Behavior* 14: 129–177.

Fuller, C., M. Simmering, G. Atinc, Y. Atinc, and B. Babin. 2016. "Common Methods Variance Detection in Business Research." *Journal of Business Research* 69, no. 8: 3192–3198.

Gagné, M., and E. L. Deci. 2005. "Self-Determination Theory and Work Motivation." *Journal of Organizational Behavior* 26, no. 4: 331–362.

Gagné, M., and J. Forest. 2008. "The Study of Compensation Systems through the Lens of Self-Determination Theory: Reconciling 35 Years of Debate." *Canadian Psychology* 49, no. 3: 225–232.

Gerhart, B., J. M. Newman, and G. T. Mikovich. 2020. *Compensation* (13th ed.). McGraw-Hill.

Geys, B., S. Connolly, H. Kassim, and Z. Murdoch. 2023. "Staff Reallocations and Employee Attitudes towards Organizational Aims: Evidence Using Longitudinal Data From the European Commission." *Public Management Review* 25, no. 12: 2323–2343. https://doi.org/10.1080/14719037.2023.2222139.

Gibbons, R., and R. Henderson. 2012. "Relational Contracts and Organizational Capabilities." *Organization Science* 23, no. 5: 1350–1364.

Gius, M. 2013. "The Effects of Merit Pay on Teacher Job Satisfaction." *Applied Economics* 45, no. 31: 4443–4451. https://doi.org/10.1080/00036846.2013.788783.

Greenberg, J. 1990. "Organizational Justice: Yesterday, Today, and Tomorrow." *Journal of Management* 16, no. 2: 399–433.

Greene, W. H. 2008. *Econometric Analysis* (6th ed.). Upper Saddle River, NJ: Prentice-Hall.

Greer, L. L. 2012. "Group Cohesion: Then and Now." *Small Group Research* 43, no. 6: 655–661.

Gruman, J. A., and A. M. Saks. 2011. "Performance Management and Employee Engagement." *Human Resource Management Review* 21, no. 2: 123–136.

Hanushek, E. A., and S. G. Rivkin. 2010. "Generalizations About Using Value-Added Measures of Teacher Quality." *American Economic Review* 100, no. 2: 267–271.

Hasnain, Z., N. Manning, and H. Pierskalla. 2012. "Performance-Related Pay in the Public Sector: A Review of Theory and Evidence." Policy Research Working Paper Series, 6043. The World Bank.

Hayes, A. F. 2012. PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling [White paper]. Available at: http://www.afhayes.com/public/process2012.pdf.

Heery, E. 1998. "A Return to Contract? Performance Related Pay in a Public Service." *Work, Employment & Society* 12, no. 1: 73–95.

Heneman, R. L., and J. M. Werner. 2005. *Merit Pay: Linking Pay to Performance in a Changing World*. Greenwich, CT: Information Age Publishing.

Hussain, I. 2012. Subjective Performance Evaluation in the Public Sector: Evidence From School Inspections. CEE Discussion Paper 135. London: Centre for the Economics of Education, London School of Economics.

Kennedy, P. 2008. *A Guide to Econometrics* (6th ed.). Malden, MA: Blackwell.

Kleibergen, F. 2007. "Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics." *Journal of Econometrics* 139, no. 1: 181–216.

Kominis, G., and C. R. Emmanuel. 2007. "The Expectancy–Valence Theory Revisited: Developing an Extended Model of Managerial Motivation." *Management Accounting Research* 18, no. 1: 49–75.

Koumenta, M. 2011. "Modernization, Privatization, and the Public Service Ethos in the United Kingdom." In *Employment in the Lean Years: Policy and Prospects for the Next Decade*, edited by D. Marsden. Chapter 14, 199.

OECD. 2005a. *Performance-Related Pay Policies for Government Employees*. Paris: OECD.

Landry, A. T., M. Gagné, J. Forest, et al. 2017. "The Relation between Financial Incentives, Motivation, and Performance: An Integrative SDT-Based Investigation." *Journal of Personnel Psychology* 16, no. 2: 61–76.

Laroche, P. 2016. "A Meta-Analysis of the Union-Job Satisfaction Relationship." *British Journal of Industrial Relations* 54, no. 4: 709–741.

Larsson, B., Y. U. Eriksson, and P. Adolfsson. 2022. "Motivating and Demotivating Effects of Performance-Related Pay in Swedish Public Sector Organizations." *Review of Public Personnel Administration* 42, no. 3: 444–463.

Latham, G. P., and C. C. Pinder. 2005. "Work Motivation Theory and Research at the Dawn of the Twenty-First Century." *Annual Review of Psychology* 56: 485–516.

Lawler, E. E., and J. L. Suttle. 1973. "Expectancy theory and job behavior." *Organizational Behavior & Human Performance* 9, no. 3: 482–503. https://doi.org/10.1016/0030-5073(73)90066-4.

Lazear, E. 1998. *Personnel Economics for Managers*. New York: Wiley.

Lazear, E. P., and P. Oyer. 2012. "Personnel Economics." In *The Handbook of Organizational Economics*, edited by R. Gibbons and J. Roberts, 188–205. Princeton University Press.

Lee, H. W. 2019. "Performance-Based Human Resource Management and Federal Employee's Roles of Goal-Clarifying Intervention, Appraisal Fairness, and Feedback Satisfaction." *Review of Public Personnel Administration* 39, no. 3: 323–348.

Levy, P. E., and J. R. Williams. 2004. "The Social Context of Performance Appraisal: A Review and Framework for the Future." *Journal of Management* 30, no. 6: 881–905.

Locke, E. A., and G. P. Latham. 2002. "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey." *American Psychologist* 57, no. 9: 705–717.

Marsden, D. 1999. *A Theory of Employment Systems: Micro-Foundations of Societal Diversity*. Oxford: Oxford University Press.

Marsden, D. 2007. "Individual Employee Voice: Renegotiation and Performance Management in Public Services." *International Journal of Human Resource Management* 18, no. 7: 1263–1278.

Marsden, D. 2010. "The Paradox of Performance-Related Pay Systems." In *Paradoxes of Modernization: Unintended Consequences of Public Policy Reform*, edited by H. Margetts and C. Hood, 185–203. Oxford: Oxford University Press.

Marsden, D. 2013. "Individual Voice in Employment Relationships: A Comparison Under Different Forms of Workplace Representation." *Industrial Relations: A Journal of Economy and Society* 52: 221–258.

Marsden, D., and R. Belfield. 2006. "Pay for Performance Where Output is Hard to Measure: the Case of Performance Pay for School Teachers." In *Advances in Industrial and Labor Relations*, edited by D. Lewin, B. Kaufman, and P. J. Gollan, Vol. 15, 1–34. Leeds: Group Publishing Limited.

Marsden, D. W., and R. Richardson. 1994. "Performing for Pay? The Effects of 'Merit Pay' on Motivation in a Public Service." *British Journal of Industrial Relations* 32, no. 2: 243–262.

McGraw, P., J. Larsen, D. Kahneman, and D. Schkade. 2010. "Comparing Gains and Losses." *Psychological Science* 21, no. 10: 1438–1445.

Murnane, R. J., and D. K. Cohen. 1986. "Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive." *Harvard Educational Review* 56, no. 1: 1–17.

Murnane, R. J., and J. P. Papay. 2010. "Teachers' Views on No Child Left Behind: Support for the Principles, Concerns About the Practices." *Journal of Economic Perspectives* 24, no. 3: 151–166.

NEU. (n.d.) Teachers' Pay Progression [Accessed in July 2023]. Available at: https://neu.org.uk/advice/your-rights-work/pay/pay-progression/teachers-pay-progression#:~:text=All%20teachers%20are%20eligible%20to,of%20the%20upper%20pay%20range.

NEU. 2017. *Pay & Pay Progression for September 2017: A Survey Report on Pay in Schools & Academies by the National Education Union*. London: National Education Union.

OECD. 2012. *Public Sector Compensation in Times of Austerity*. Paris: OECD.

Ofsted. 2014. The Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2013/14. Manchester: Office for Standards in Education, Children's Services and Skills (Ofsted).

Ofsted. 2015. School Inspection Handbook: Handbook for Inspecting Schools in England under Section 5 of the Education Act 2005 (as amended by the Education Act 2011). Manchester: Office for Standards in Education, Children's Services and Skills (Ofsted).

Ogbonnaya, C., K. Daniels, and K. Nielsen. 2017. "Does Contingent Pay Encourage Positive Employee Attitudes and Intensify Work?" *Human Resource Management Journal* 27, no. 1: 94–112.

Oyer, P., and S. Schaefer. 2010. Personnel Economics: Hiring and Incentives. National Bureau of Economic Research, Working Paper 15977. Washington: NBER.

Pham, L. D., T. D. Nguyen, and M. G. Sringer. 2021. "Teacher Merit Pay: A Meta-Analysis." *American Educational Research Journal* 58, no. 3: 527–566.

Pignata, S., A. H. Winefield, C. Provis, and C. M. Boyd. 2016. "A Longitudinal Study of the Predictors of Perceived Procedural Justice in Australian University Staff." *Frontiers in Psychology* 25, no. 7: 1271. https://doi.org/10.3389/fpsyg.2016.01271.

Podgursky, M. J., and M. G. Springer. 2007. "Teacher Performance Pay: A Review." *Journal of Policy Analysis and Management* 26, no. 4: 909–949.

Podsakoff, P. M., S. B. MacKenzie, J. Y. Lee, and N. P. Podsakoff. 2003. "Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies." *Journal of Applied Psychology* 88, no. 5: 879–903.

Podsakoff, P. M., S. B. MacKenzie, and N. P. Podsakoff. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It." *Annual Review of Psychology* 63: 539–569.

Podsakoff, P. M., B. P. Niehoff, S. B. MacKenzie, and M. L. Williams. 1993. "Do Substitutes for Leadership Really Substitute for Leadership? An Empirical Examination of Kerr and Jermier's Situational Leadership Model." *Organizational Behavior and Human Decision Processes* 54, no. 1: 1–44.

Propper, C., and D. Wilson. 2003. "The Use and Usefulness of Performance Measures in the Public Sector." *Oxford Review of Economic Policy* 19, no. 2: 250–268.

Richardson, R. 1999. *Performance Related Pay in Schools: An Assessment of the Green Papers. A Report for the National Union of Teachers*. London: National Union of Teachers.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73, no. 2: 417–458.

Rockoff, J. E., and C. Speroni. 2010. "Subjective and Objective Evaluations of Teacher Effectiveness." *American Economic Review* 100, no. 2: 261–266.

Rynes, S. L., B. Gerhart, and L. Parks. 2005. "Personnel Psychology: Performance Evaluation and Pay for Performance." *Annual Review of Psychology* 56: 571–600.

Sargan, J. D. 1958. "The Estimation of Economic Relationships Using Instrumental Variables." *Econometrica: Journal of the Econometric Society* 26: 393–415.

Shadman-Pajouh, M. 2023. "Management by Objectives." In *Sage Business Foundations: Human Resource Management*, edited by H. El-Farr, 23–41. SAGE Publications, Inc. https://doi.org/10.4135/9781071924181.

Shapiro, D. L., and B. L. Kirkman. 2001. "Anticipatory Injustice: the Consequences of Expecting Injustice in the Workplace." In *Advances in Organizational Justice*, edited by J. Greenberg, and R. Cropanzano, Chapter 5, 152–178. Standford, CA: Stanford University Press.

Shrout, P. E., and J. L. Fleiss. 1979. "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychological Bulletin* 86, no. 2: 420–428.

Staiger, D., and J. Stock. 1997. "Instrumental Variables Regression With Weak Instruments." *Econometrica* 65: 557–586.

Staiger, D. O., and E. R. Jonah. 2010. "Searching for Effective Teachers With Imperfect Information." *Journal of Economic Perspectives* 24, no. 3: 97–118.

Stock, J., and M. Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by W. Donald, K. Andrews, and J. H. Stock, Chapter 5, 80–108. Cambridge: Cambridge University Press.

STRB. 2012. *School Teachers Pay Review Body: Twenty-First Report*. London: Stationery Office.

STRB. 2023. *School Teachers Pay Review Body: Thirty-Third Report*. London: Stationery Office.

STRB. 2024. *School Teachers Pay Review Body: Thirty-Fourth Report*. London: Stationery Office.

Taylor, E., and J. H. Tyler. 2011. The Effect of Evaluation on Performance: Evidence From Longitudinal Student Achievement Data of Mid-Career Teachers. NBER Working Paper No. 16877. Cambridge, MA: National Bureau of Economic Research.

Taylor, M. S., K. B. Tracy, M. K. Renard, J. K. Harrison, and S. J. Carroll. 1995. "Due Process in Performance Appraisal: A Quasi-Experiment in Procedural Justice." *Administrative Science Quarterly* 40, no. 3: 495–523.

Thurston Jr., P. W., and L. McNall. 2010. "Justice Perceptions of Performance Appraisal Practices." *Journal of Managerial Psychology* 25, no. 3: 201–228.

Wenzel, A. K., T. Krause, and D. Vogel. 2019. "Making Performance Pay Work: the Impact of Transparency, Participation, and Fairness on Controlling Perception and Intrinsic Motivation." *Review of Public Personnel Administration* 39: 232–255.

Wilkinson, A., T. Dundon, J. Donaghey, and R. B. Freeman. 2014. "Employee Voice: Charting New Terrain." In *Handbook of Research on Employee Voice*, edited by A. Wilkinson, T. Dundon, J. Donaghey, and R. B. Freeman, Chapter 1, 1–19. Cheltenham: Edward Elgar.

Wooldridge. 2015. "Instrumental Variables and Two Stage Least Squares in Wooldridge." In *Introductory Econometrics* (4th ed.).

Supporting Information

Additional supporting information can be found online in the Supporting Information section.