The dilemma of performance appraisal

Peter Prowse and Julie Prowse

Summary
Purpose – This paper aims to evaluate the aims and methods of appraisal, and the difficulties encountered in the appraisal process.
Design/methodology/approach – The paper uses a review of the literature to evaluate the development of appraisals and argues that the critical area of line management development that was identified as a critical success factor in appraisals has been ignored in the later literature evaluating the effectiveness of performance appraisals.
Findings – The review identifies the lack of theoretical development in appraisal and argues the psychological approach of analysis and a more critical realisation of appraisal, re-evaluating the challenge to remove subjectivity and bias in judgement of appraisal.
Research limitations/implications – Further evaluation of key interpersonal skills is required for appraisal systems to develop performance.
Practical implications – The use and design of performance pay in public and private services linked to appraisal have not always improved organisational performance and can contribute to reduced motivation.
Originality/value – Little research has evaluated the current increases in using appraisals and the changes in focus from appraisal to performance management. The paper adds value to the existing body of knowledge and offers insights for practitioners and researchers.
Keywords Performance appraisal, Performance measurement (quality), Performance related pay
Paper type Literature review

Introduction
This article will define and outline performance management and appraisal. It will start by evaluating what form of performance is evaluated, then establish links to the development of different performance traditions (psychological tradition, management by objectives, motivation and development). It will outline the historical development of performance management then evaluate high performance strategies using performance appraisal. It will then evaluate the continuing issue of subjectivity and ethical dilemmas regarding the measurement and assessment of performance. The article then examines how organisations measure performance before evaluating the research on recent trends in performance appraisal.

Objectives
This article evaluates the historical development of performance appraisal from management-by-objectives (MBO) literature before evaluating the debates on the linkages between performance management and appraisal. It will outline the use of performance management as a method for managing individual performance in organisations. The techniques used to increase organisational commitment and job...
satisfaction will be critically evaluated. It will also examine the transatlantic debates on efficiency and effectiveness in the North American and UK literature and the contribution of performance appraisal to individual and organisational performance.

**Use of performance appraisal**

Performance appraisal is potentially a key tool for organisations to make the most of their human resources and the use of appraisal is widespread. An estimated 80-90 per cent of organisations in the USA and UK use appraisals, and there has been an increase from 69 per cent to 87 per cent of organisations between 1998 and 2004 using formal performance management systems (Armstrong and Baron, 2005).

Between 1998 and 2004 a survey from the Chartered Institute of Personnel and Development of 562 firms found 506 were using performance appraisal in UK (Chartered Institute of Personnel and Development, 2007). What is also vital to emphasise is the rising use of performance appraisal feedback beyond professionals and managers to nearly 95 per cent of workplaces as the WERS (2004) survey (Kersley et al., 2006). The development and extension of appraisals to a large proportion of the UK workforce and the coverage of non-managerial occupations has extended the use of appraisals in private and public sectors. However, there is little evidence or evaluation of the effectiveness of appraisal.

**The purpose of appraisals**

What is the purpose of appraisals, how it is researched and used in practice in organisations? Randell (1994) argues that they are the systematic evaluation of individual performance linked to workplace behaviour and/or specific criteria. Appraisals often take the form of an appraisal interview, usually annual, supported by standardised forms/paperwork. The key objective of appraisal is to provide employees with feedback on their performance provided by the line manager. There are three key questions for quality of feedback:

1. What and how are observations on performance made?
2. Why and how are they discussed?
3. What determines the level of performance in the job?

It has been argued that this process cannot be performed effectively unless the line manager or person providing feedback has the interpersonal interviewing skills to provide that feedback to people being appraised. This has been defined as the ‘Bradford Approach’, which places a high priority on appraisal skills development (Randell, 1994). This approach identifies the linkages between involving, developing, rewarding and valuing people at work.

**Historical development of appraisal**

The historical development of performance feedback has developed from a range of approaches. Formal observations of individual work performance were reported in Robert Owens's factory in New Lanark in the early 1800s, hanging a piece of coloured wood over machines to indicate the superintendent's assessment of the previous day's conduct (white for excellent, yellow, blue and then black for poor performance) (Cole, 1925). The twentieth century led to F.W. Taylor and his measured performance and the scientific management movement (Taylor, 1964). The psychological tradition developed in the 1930s used approaches that identified personality and performance used feedback from graphic rating scales, a mixed standard of performance scales noting behaviour in Likert-scale ratings, providing evidence to recruit and identify management potential in the field of selection. Later developments removed the middle scale from a five-point scale to develop into forced-choice scale judgements to avoid central ratings. The evaluation also included narrative statements and comments to support the ratings (Mair, 1958).
In the 1940s behavioural methods were developed using a motivational approach. These included behavioural anchored rating scales (BARS), behavioural observation scales (BOS), behavioural evaluation scales (BES), critical incident, and job simulation. All these judgements were used to determine the specific levels of performance criteria to specific issues such as customer service and rated in factors such as “excellent”, “average” or “needs to improve” or “poor”. These feedback ratings are assigned numerical values and added to a statement or narrative comment by the assessor but were essentially developmental and also led to identification of any potential need for training and, more importantly, to identify talent for careers in line management supervision and future managerial potential.

Post-1945 developed into the results-oriented approaches and led to the development of management by objectives (MBO). This provided aims and specific targets to be achieved and within time frames such as specific sales, profitability and deadlines, with feedback on previous performance (Wherry, 1957). The deadlines may have required alteration and led to specific performance rankings of staff. It also provided a forced distribution of rankings of comparative performance and paired comparison ranking of performance and setting and achieving objectives.

In the 1960s the development of self-appraisal by discussion led to specific time and opportunity for the appraisee to evaluate their performance reflectively in the discussion and the interview developed into a conversation on a range of topics that the appraisee needed to discuss in the interview. Until this period the success of the appraisal was dependent on the skill of interviewer. In the 1990s 360-degree appraisal developed, where information was sought from a wider range of sources and the feedback was no longer dependent on the manager-subordinate power relationship but included groups appraising the performance of line managers and peer feedback from peer groups on individual performance (Redman and Snape, 1992). The final development of appraisal interviews developed in the 1990s with the emphasis on linking performance with financial reward, which will be discussed later in the article.

Measures of performance

The dilemma of appraisal has always been developing performance measures, and the use of appraisal communicating individual performance is the key part of this process. Quantitative measures of performance communicated as standards in the business and industry level standards are to individual performance. The introduction of techniques such as the Balanced Scorecard developed by Kaplan and Norton (1992) extended performance measures and evaluation included financial, customer evaluation, feedback on internal processes and learning and growth. Performance standards also included qualitative measures that argue that there is an over-emphasis on the metrics of the quantitative approach above the definitions of quality services and total quality management. In terms of performance measures there was a transformation in the literature and a move in the 1990s to financial rewards linked to the level of performance. This will be discussed later in the article.

Criticism of appraisals

Critiques of appraisal have continued as appraisals have increased in use and scope across sectors and occupations. The dominant critique is the management framework using appraisal as an “orthodox” technique that seeks to remedy the weakness and propose of appraisals as a system to develop performance (Bach, 2005). This “orthodox” approach argues there are conflicting purposes of appraisal (Strebler et al., 2001). Appraisal can motivate staff by clarifying objectives and setting clear future objectives with provision for training and development needs to establish the performance objective. These conflict with assessing past performance and distribution of rewards based on past performance (Bach, 2005, p. 301). Employees are reluctant to confide any limitations to and concerns with their current performance as this could impact on their merit-related reward or promotion
opportunities (Newton and Findley, 1996, p. 43). This conflicts with performance appraisal as a developmental process as appraisers are challenged with differing roles as both monitors and judges of performance, and an understanding counsellor, which Randell (1994) argues few managers receive the training to perform. Managerial reluctance to criticise also stems from classic evidence from McGregor that they are reluctant to make negative judgements on an individual's performance as it could be de-motivating, leading to appraisee accusations of lack of managerial support and contribution to an individual's poor performance (McGregor, 1957). One consequence of conflict avoidance is managers rate all criterion in the middle rating point, known as the “central tendency”. In a study of senior managers by Longenecker et al. (1987), it was found that organisational politics influenced the ratings of 60 senior executives. Politics involved deliberate attempts by individuals to enhance or protect self-interests when conflicting courses of action are possible, and ratings and decisions were affected by potential sources of bias or inaccuracy in their appraisal ratings (Longenecker and Ludvig, 1990). There are methods of further bias beyond Longenecker’s evidence. Political judgements were distorted further by over-rating some clear competencies in performance rather than being critical across the range of measures, known as the “halo effect”, and if some competencies are lower they may prejudice the judgment across the positive reviews, known as the “horns effect” (Advisory, Conciliation and Arbitration Service, 1996).

Some ratings may only include recent events, and these are known as “recency effects”. In this case only recent events are noted compared to managers gathering and using evidence throughout the appraisal period. A particular concern is the consistency and equity of appraisal ratings, which may be distorted by gender, ethnicity and the ratings of appraisers themselves. Studies in both the USA and the UK have highlighted subjectivity in terms of gender (Alimo-Metcalf, 1991; White, 1999) and ethnicity of the appraisee and appraiser (Geddes and Konrad, 2003). Suggestions and solutions on resolving bias will be reviewed later.

The second analysis is the “radical critique” of appraisal. This critical management literature argues that appraisal and performance management are more covert forms of management control (Newton and Findley, 1996, Townley, 1993). This argues that tighter management control over employee behaviour can be achieved by the extension of appraisal to both manual and professional workers. This develops the literature of Foucault (1979) using power and surveillance and evidence uses cases in examples of public service control on professionals such as teachers (Healy, 1997) and university professionals (Townley, 1990). This evidence argues the increased control of public services using appraisal as a method of control and the outcome of managerial objectives ignoring the developmental role of appraisal with ratings awarded for people who accept and embrace the culture and organisational values. This literature ignores employee resistance and the use of professional unions to challenge attempts to exert control over professionals and staff in the appraisal process (Bach, 2005, p. 306).

One of the different issues of removing bias was the use of the “test metaphor” (Folger et al., 1992). This was based on the assumption that appraisal ratings were a technical question of assessing “true” performance and there needed to be increased reliability and validity of appraisal as an instrument to develop motivation and performance. The sources of rater bias and errors can be resolved by improved organisational justice and increasing the reliability of the appraiser’s judgement. However, there were problems, such as an assumption that one can state job requirements clearly, that the organisation can be “rational” with objectives that reflect values, and that judgments by appraisers are value-free from political agendas and personal objectives. Secondly there is the issue of subjectivity, where decisions on appraisal are rated by a “political metaphor” (Hartle, 1992). This metaphor argues that appraisal is often done badly because there is a lack of training for appraisers and appraisers may see the appraisal as being of no value. Organisations in this context are ‘political’ and appraisers seek to maintain performance from subordinates and view appraisals as internal customers to satisfy. This context forces managers using appraisal to avoid interpersonal conflict and develop strategies for their own personal advancement and
avoiding censure from higher managers. In this context, appraisal ratings become political judgements and managers seek to avoid interpersonal conflict. The approaches of “test” and “political” metaphors of appraisal are both inaccurate, lacking objectivity and poor judgement of employee performance. The question is: how can organisations resolve this lack of objectivity?

Resolving lack of objectivity for appraisal ratings

Grint (1993) argues that the solutions to improve objectivity lie in part with McGregor’s (1957) classic critique by retraining and the removal of “top down” ratings by managers and their replacement with multiple-rater evaluation using 360-degree appraisal. This attempts to resolve bias and objectivity by upward performance appraisal. The solution of multiple-rater reporting uses internal colleagues, customers and recipients of services and will reduce the subjectivity and inequity of appraisal ratings. This solution may be influenced further by the rise in the need to evaluate project teams and increasing levels of teamwork to include peer assessment. The solutions also in theory mean increased closer contact with individual managers and appraisees and increasing services linked to customer evaluations. A manager in 360-degree appraisal collates feedback rather than judges performance and summarises evaluations. The validity of upward appraisal means the removal of subjective appraisal ratings. This approach is also suggested to resolve gender bias against women in appraisal ratings (Fletcher, 1999). However, negative feedback still demotivates, and managers’ role in detailed feedback still requires skills and sensitivity in discussions. There are concerns with the accuracy of appraisal objectivity, as Walker and Smither’s (1999) five-year study of 252 managers still identified issues with subjective ratings in 360-degree appraisals. There are still issues on the subjectivity of appraisals beyond the solutions of appraisal skills training.

The contribution of appraisal is strongly related to employee attitudes and strong association with job satisfaction (Fletcher and Williams, 1996). The evidence on appraisal still remains positive in terms of reinvigorating social relationships at work (Townley, 1993) and one benefit of the widespread adoption in large public services in the UK, such as the National Health Service (NHS), is the valuable contribution to line managers’ discussion with staff on their past performance, discussing personal development plans and training and development as positive issues (Redman and Snape, 1992). One final concern is appraisal related to employee reward, which we now discuss.

Linking appraisals with reward management

Appraisal and performance management have been inextricably linked to employee reward since the development of strategic human resource management in the 1980s. The early literature on appraisal linked appraisal with employee control (Randell, 1994; Grint, 1993; Townley, 1993, 1999) and discussed the use of performance-related reward in appraisal. However, the literature in the 1990s substituted the chapter titles of employee “appraisal” with “performance management” (Bach, 2005; Storey, 2007) and focussed on performance and performance pay. The appraisal and performance pay link has developed into debates to three key issues:

1. Has performance pay related to appraisal grown in use?
2. What type of performance do we reward?
3. Who judges management standards?

The first discussion on influences of growth of performance pay schemes is the assumption that increasing the linkage between individual effort and financial reward increases levels of performance. This assumption is increasing used in schemes in both the public and private sectors (Bevan and Thompson, 1992; Armstrong and Baron, 2005). The drive to increase public-sector performance effort and setting of targets may even be inconsistent in the experiences of some organizational settings aimed at achieving long-term targets (Kessler and Purcell, 1992; Marsden, 2007). The development of merit-based pay based on
performance assessed by a manager is rising in the UK. Marsden (2007) reported that the increased use of performance appraisals as a basis for merit pay “are used in 65 per cent of public sector and 69 per cent of the private sector employees where appraisal covered all non-managerial staff” (p. 109). Merit pay has also grown in use, with 20 per cent of workplaces using performance related schemes in 1998 compared to 32 per cent in the same organisations in 2004 (Kersley et al., 2006, p. 191). The achievement of satisfactory ratings or above satisfactory performance averages have been used as evidence to reward individual performance ratings in the UK civil service (Marsden, 2007).

The second issue is what forms of performance are rewarded. The use of past appraisal ratings as evidence of achieving merit-related payments linked to achieving higher performance was the predominant factor developed in the public services. The evidence on setting performance targets have been, as Bach and Sisson (2000, p. 280) reported, “inconsistent within organizations and problematic for certain professional or less skilled occupations where goals have not been easily formulated”. There has been inconclusive evidence from organisations on the impact of performance pay and its effectiveness in improving performance. Evidence from a number of individual performance pay schemes report organisations suspending or reviewing them on the grounds that individual performance reward has produced no effect in performance or has even demotivated staff (Bach and Sisson, 2000, p. 281). More in-depth studies setting performance goals followed by appraisal reported reduced motivation whilst maintaining productivity, achieved by managers imposing increased performance standards (Marsden and Richardson, 1994). As Randell (1994) highlighted earlier, the potential objectivity and self-criticism in appraisal reviews become areas that appraisees refuse to acknowledge with appraisers if this leads to a reduction in their merit pay. Objectivity and self-reflection for development becomes a weakness that appraisees fail to accept as a developmental issue if any self-evaluation potentially reduces their merit reward. Reviews of civil service merit pay in four major UK civil service agencies and the National Health Service concluded that existing forms of performance pay and performance management had failed to motivate many staff (Makinson, 2000; Marsden, 2007). The conclusions were that that employees found individual performance pay divisive and led to reduced willingness to cooperate with management, citing managerial favourites and manipulation of appraisal scores to lower ratings to avoid paying incentive rewards (Makinson and French, 1998). This has clear implications on the relationship between line managers and appraisees and the de-motivational consequences reduce commitment, providing clear evidence of the danger of linking individual performance appraisal to reward in the public services. Employees focus on the issues that gain key performance focus by focusing on specific objectives related to key performance indicators rather than all personal objectives. A study of banking performance pay highlighted that centrally imposed targets were unattainable and showed the existence of a range of 20 performance targets with narrow short-term financially oriented goals (Lewis, 1998). The narrow focus on key targets and neglect of other performance aspects leads to tasks not being delivered.

This final issue of judging management standards has already highlighted issues of inequity and bias based on gender (Beyer, 1990; Chen and DiTomaso, 1996; Fletcher, 1999). The solutions suggested to resolve discrimination include enhanced interpersonal skills training, and the increased equitable use of 360-degree appraisal as a method to evaluate feedback from colleagues, as this reduces the use of the “political metaphor” (Randell, 1994; Fletcher, 1999).

Measures linking performance to service improvement require a wider approach to enhanced work design and motivation to develop and enhance employee job satisfaction. The design of appraisal systems has identified the need for improved linkages between effort and performance in the private sector and increasing feedback and awareness in the public sector (Fletcher and Williams, 1996, p. 176). Where rises in pay were determined by achieving critical rated appraisal objectives, employees were less self-critical and open to any developmental needs in a performance review.
Conclusion
As performance appraisal provides a major potential for employee feedback that could link strongly to increasing motivation, an opportunity to clarify goals and achieve long-term individual performance and career development, why does it still suffer from what Randell (1994) describes as a muddle and confusion which still surrounds the theory and practice? There are key issues that require resolution and a great deal depends on the extent to which one has a good relationship with one’s line managers. Barlow (1989) argued: “if you get off badly with your first two managers, you may just as well forget it” (p. 515). The evidence on the continued practice of appraisals is that they are still institutionally elaborated systems of management appraisal and their development is a significant rhetoric in the apparatus of bureaucratic control by managers (Barlow, 1989). In reality companies create, review, change and even abolish appraisals if they fail to develop and enhance organisational performance (Kessler, 2000). Despite all the criticism and evidence, the critics have failed to suggest an alternative for a process that can provide feedback, develop motivation, identify training needs and potential and evidence that can justify potential career development and justify reward (Hartle, 1997).

References


**Further reading**


**About the authors**

Peter Prowse is a Senior Lecturer in Human Resource Management and Employee Relations at University of Bradford School of Management. His research interests are public services reform and management.

Julie Prowse is a Senior Lecturer at University of Bradford School of Health Studies. Her research interests include National Health Service reforms and the implications for employees.