

Technical report

December 2016

Rapid evidence assessment of the
research literature on the effect of
performance appraisal
on workplace performance



The CIPD is the professional body for HR and people development. The not-for-profit organisation champions better work and working lives and has been setting the benchmark for excellence in people and organisation development for more than 100 years. It has more than 140,000 members across the world, provides thought leadership through independent research on the world of work, and offers professional training and accreditation for those working in HR and learning and development.

Rapid evidence assessment of the research literature on the effect of performance appraisal on workplace performance

Technical report

Contents

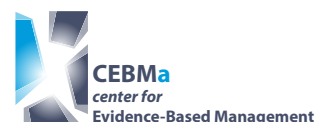
Foreword	2
Introduction	3
1 Methodology	4
2 Findings	6
3 Synthesis	16
Conclusion	18
Limitations	19
Endnotes	20
References	21
Appendix 1	26
Appendix 2	27

Acknowledgements

We would like to thank the team behind this research at the Center for Evidence-Based Management (CEBMa). The report was written by Dr Eric Barends, Barbara Janssen and Pietro Marengo, with the support of Professor Rob Briner and Professor Denise Rousseau, all of CEBMa.

About CEBMa

The Center for Evidence-Based Management is a non-profit member organisation dedicated to promoting evidence-based practice in the field of management. It provides support and resources to managers, consultants, organisations, teachers, academics and others interested in learning more about evidence-based management.



Foreword

As a professional body, the CIPD helps HR professionals and functions to develop effective strategies and practices in people management. We believe that applied research is a crucial step to achieving this. We thus see an important part of our role as making quality research available, distilling it into accessible forms and drawing out practical implications.

This technical report presents the methods and findings of a rapid evidence assessment (REA), a truncated form of systematic review, on the topic of performance appraisal. It is accompanied by another technical report of an REA on goal setting (Barends et al 2016). The insight and implications of both technical reports, which are written by the Center for Evidence-Based Management (CEBMA), are discussed in the discussion report, *Could do Better: Assessing what works in performance management* (Gifford 2016).¹

In publishing this technical report, we provide a step-by-step account of the evidence our REA uncovered on what is meant by performance appraisal, how it is assumed to work and what influences its effectiveness. A number of the academic papers referenced are accessible through the EBSCO online journals portal for CIPD members.²

We hope this report provides a useful reference and pointer to further reading on this important aspect of performance management.

Jonny Gifford

Adviser, Organisational Behaviour
CIPD

Introduction

This technical report presents the methods and findings of a rapid evidence assessment (REA), a truncated form of systematic review, on the topic of performance appraisal. It is accompanied by another technical report of an REA on goal setting. The insight and implications of both technical reports are discussed in the discussion report, *Could do Better: Assessing what works in performance management*. All reports are available at cipd.co.uk/coulddobetter

Rationale for this review

Despite the relevance of performance ratings within the domain of human resource management, both academics and practitioners have always had a somewhat uneasy relationship with them. Some academics question whether performance appraisals provide meaningful information, whereas others have even suggested that undertaking such reviews should be discontinued entirely (Hoffman et al 2012). Given the widespread use of performance appraisals within management practice, the CIPD approached the Center for Evidence-Based Management (CEBMA) to undertake a review to understand what is known in the scientific literature about the reliability and validity of performance appraisal and the way in which this may impact workplace performance. This review will present an overview of this evidence.

Main question: What does the review answer?

What is known in the scientific literature about the impact of performance appraisal on workplace performance?

Supplementary questions

Other issues raised, which will form the basis of our conclusion to the main question above, are:

- 1 What is meant by performance appraisal? (What is it?)
- 2 What is the assumed causal mechanism? (How is it supposed to work?)
- 3 What is the effect of performance appraisal on workplace performance?
- 4 What is known about possible moderators and/or mediators that affect the relationship between performance appraisal and workplace performance?
- 5 What is known about the reliability and validity of performance appraisal?

1 Methodology

Search strategy: How was the research evidence sought?

The following four databases were used to identify studies: ABI/INFORM Global, Business Source Premier, PsycINFO and Web of Science. The following generic search filters were applied to all databases during the search:

- 1 scholarly journals, peer-reviewed
- 2 published 1980–2016 for meta-analyses, and 2000–16 for primary studies
- 3 articles in English.

A search was conducted using combinations of different search terms, such as ‘performance appraisal’, ‘performance review’, ‘performance evaluation’, ‘annual review’ and ‘employee evaluation’. In addition, the references listed in the studies retrieved were screened in order to identify additional articles for possible inclusion in the REA.

We conducted 17 different search queries and screened the titles and abstracts of more than 250 studies. An overview of all search terms and queries is provided in Appendix 1.

Selection process: How were the studies selected?

Two reviewers worked independently to identify which studies should be included. The inter-rater agreement was 92.5%. Where the reviewers disagreed on selection, a third reviewer – with no prior knowledge of the initial reviewers’ assessments – assessed whether the study was appropriate

for inclusion. The decision of the third reviewer was final.

Selection took place in two phases. First, the titles and abstracts of the 250+ studies identified were screened for their relevance to this review. In case of doubt or lack of information, the study was included. Duplicate publications were removed. This first phase yielded 41 secondary studies (meta-analyses) and 48 primary studies.

Second, studies were selected based on the full text of the article according to the following inclusion criteria:

- 1 **type of studies:** quantitative, empirical studies
- 2 **measurement:** (a) studies in which the effect of performance appraisal on organisational outcomes was measured, or (b) studies in which the effect of moderators and/or mediators on performance appraisal was measured
- 3 **context:** studies related to workplace settings
- 4 **level of trustworthiness:** studies that were graded level C or above.

In some cases where influential studies were referenced that had not been identified in our search because of the search terms used, we included these additional studies as ‘bycatch’.

This second phase yielded 23 secondary studies and 37 primary studies. An overview of the selection process is provided in Appendix 2.

Critical appraisal: What is the quality of the studies included?

In almost any situation it is possible to find a scientific study to support or refute a theory or a claim, and sometimes to quite a large degree. It is therefore important to determine which studies are trustworthy (that is, valid and reliable) and which are not. The trustworthiness of a scientific study is first determined by its methodological appropriateness.

For cause-and-effect claims (that is, if we do A, will it result in B?), a study has a high methodological appropriateness when it fulfils the three conditions required for causal inference: co-variation, time-order relationship, and elimination of plausible alternative causes (Shaughnessy and Zechmeister 1985). A study that uses a control group, random assignment and a before-and-after measurement is therefore regarded as the ‘gold standard’.³ Non-randomised studies and before-after studies come next in terms of appropriateness. Cross-sectional studies (surveys) and case studies are regarded as having the greatest chance of showing bias in the outcome and therefore sit lower down in the ranking in terms of appropriateness. Meta-analyses in which statistical analysis techniques are used to pool the results of controlled studies are therefore regarded as the most appropriate design.

To determine the methodological appropriateness of the research

design of the studies included, the classification system of Shadish et al (2002) and Petticrew and Roberts (2006) was used. The four levels of appropriateness used for the classification are shown in Table 1.

It should be noted, however, that the level of methodological appropriateness as explained in Table 1 is relevant only in assessing the validity of a cause-and-effect relationship that might exist between an intervention (for example performance appraisal) and its outcomes (performance), which is the purpose of this review. A case study, for instance, is a strong design for assessing why an effect has occurred or how an intervention might be (un)suitable in a particular context; it does a poor job of assessing the existence or strength of a cause-and-effect relationship (Donnelly and Trochim 2007).

In addition, a study's trustworthiness is determined by its methodological quality (its

strengths and weaknesses). For instance, was the sample size large enough and were reliable measurement methods used? To determine methodological quality, all the studies included were systematically assessed on explicit quality criteria. Based on a tally of the number of weaknesses, the trustworthiness was downgraded and the final level was determined as follows: a downgrade of one level if two weaknesses were identified; a downgrade of two levels if four weaknesses were identified, and so on.

Finally, the effect sizes were identified. An effect (for example a correlation, Cohen's d or omega) can be statistically significant but may not necessarily be of practical relevance: even a trivial effect can be statistically significant if the sample size is big enough. For this reason, the effect size – a standard measure of the magnitude of the effect – of the studies included was assessed. To determine the magnitude of an effect, Cohen's rules of thumb (Cohen 1988) were

applied. According to Cohen, a 'small' effect is an effect that is visible only through careful examination. A 'medium' effect, however, is one that is 'visible to the naked eye of the careful observer'. Finally, a 'large' effect is one that anybody can easily see because it is substantial.

Outcome of the critical appraisal

The overall quality of the studies included was moderate to high. Most of the 23 secondary studies were based on cross-sectional studies and were therefore graded level B or lower, with only seven qualified as level A. Of the 37 primary studies, 20 qualified as randomised controlled studies and were therefore graded level A. The remaining 17 studies concerned quasi-experimental or longitudinal designs and were graded level B or lower.

Table 1: Four levels of appropriateness used for classification

Design	Level
Systematic review or meta-analysis of randomised controlled studies ⁴	AA
Systematic review or meta-analysis of non-randomised controlled before–after studies	A
Randomised controlled study	
Systematic review or meta-analysis of non-randomised controlled or before–after studies	B
Non-randomised controlled before–after study	
Interrupted time series	
Systematic review or meta-analysis of cross-sectional studies	C
Controlled study without a pre-test or uncontrolled study with a pre-test	

2 Findings

Figure 1: Links in the performance appraisal chain



‘Research on performance appraisal dates back at least as far as the early 1920s and has continued to the present day.’

Question 1: What is meant by performance appraisal?

Performance appraisal is one of the most widely studied topics in the domain of management. Research on performance appraisal dates back at least as far as the early 1920s and has continued to the present day. A search in ABI/INFORM on the term ‘performance appraisal’ in the title or abstract yields more than 1,200 results of peer-reviewed papers published in scholarly journals, spanning a period of six decades. One of the earliest academic papers that explicitly uses the term performance appraisal is ‘Appraisal of Job Performance’ by Stephen Halbe, published in 1951. Since then many definitions have been put forward. One of the most widely used definitions is provided by Griffin and Ebert (2004, who describe performance appraisal as the *‘formal evaluation of an employee’s job performance in order to determine the degree to which the employee is performing effectively’* (p216). Other definitions point out that it is typically an evaluation process in which quantitative scores based on predetermined criteria are assigned and shared with the employee being evaluated (for example, DeNisi and Pritchard 2006). In addition, most authors emphasise that performance

appraisal is a process composed of several elements. When examined closely, most definitions seem to have the common elements (see Figure 1):

- past performance
- establishing goals/objectives
- rating based on predetermined criteria
- judgement
- formal feedback of judgement
- future performance.

Although the primary purpose of such an appraisal is to enhance the performance or productivity of employees (and thus the organisation), most organisations use them for either administrative or developmental reasons. Developmental performance appraisals are used to identify an employee’s strengths and weaknesses and their training needs, whereas performance appraisals for administrative reasons are used to decide on salary and promotion issues, to validate selection criteria, to decide on termination of contracts and redundancies, or to meet legal requirements.

Question 2: What is the assumed causal mechanism? (How is it supposed to work?)

The assumed causal mechanism of performance appraisal is based on three theories: social comparison theory (Festinger 1954), feedback intervention theory (Kluger and DeNisi 1996), and equity theory (Adams 1965).

Social comparison theory suggests that individuals tend to compare themselves with others to make judgements regarding their performance. They are concerned not only about their performance in an absolute sense, but also about how they measure up in relation to relevant peers. In addition, this theory posits that individuals have a strong desire to improve their performance when faced with unfavourable comparative information.

Feedback intervention theory suggests that when confronted with a discrepancy between what they wish to achieve and the feedback received, individuals are strongly motivated to attain a higher level of performance.⁵ The practice of performance appraisal therefore assumes that informing an employee about the discrepancies between the organisation's standard and their current performance – implying that they are achieving lower than most other colleagues – will motivate the employee to achieve a higher level of performance.

Finally, equity theory states that employees compare themselves with each other in terms of input and outcomes (Walster et al 1978). High-performers, seeing that poor performers get lower appraisal scores – and, as a consequence, receive lower rewards – might feel that an equitable balance is being established and be motivated to continue their high-quality work,

whereas underperformers are motivated to put in more effort to achieve on a higher level.

Question 3: What is the effect of performance appraisal on workplace performance?

To measure the effect of performance appraisal on workplace performance would require an evaluation of a large number of populations and contexts where performance appraisal was applied, and the measurement of a wide range of performance outcomes, preferably by means of a meta-analysis of a large number of double-blind, randomised controlled studies. Such studies do not exist, and might well be too difficult to carry out.

However, there is wide consensus among both scholars and practitioners that performance appraisal, in general, can have a positive impact on a wide range of organisational outcomes, such as task performance, productivity, organisational citizenship behaviour, satisfaction and commitment. As stated above, both social comparison theory and feedback theory posit that providing feedback to employees regarding their relative performance can enhance employee productivity.⁶

The scientific literature on feedback performance interventions, however, suggests a caveat. Several researchers have pointed out that feedback may not always be effective. In fact, several meta-analyses have demonstrated that feedback interventions have highly variable effects on performance – in some situations, feedback improves performance, but in other situations it has no apparent effect or even harms it (Kluger and DeNisi 1996, Smither et al 2005). This finding

suggests that the relationship between performance appraisal and performance outcome is often complicated and is contingent upon a large number of moderators and mediators.⁷ As a consequence, the key question is not, 'What is the effect of performance appraisal on workplace performance?', but 'Given the target group, the objectives and the context involved, what are the factors moderating or mediating the effect of performance appraisal that need to be taken into account?'

Figure 2: 'Distribution (histogram) of 607 effects (ds) of feedback intervention on performance' (Kluger and DeNisi 1996)
 Adapted with permission from the American Psychological Association

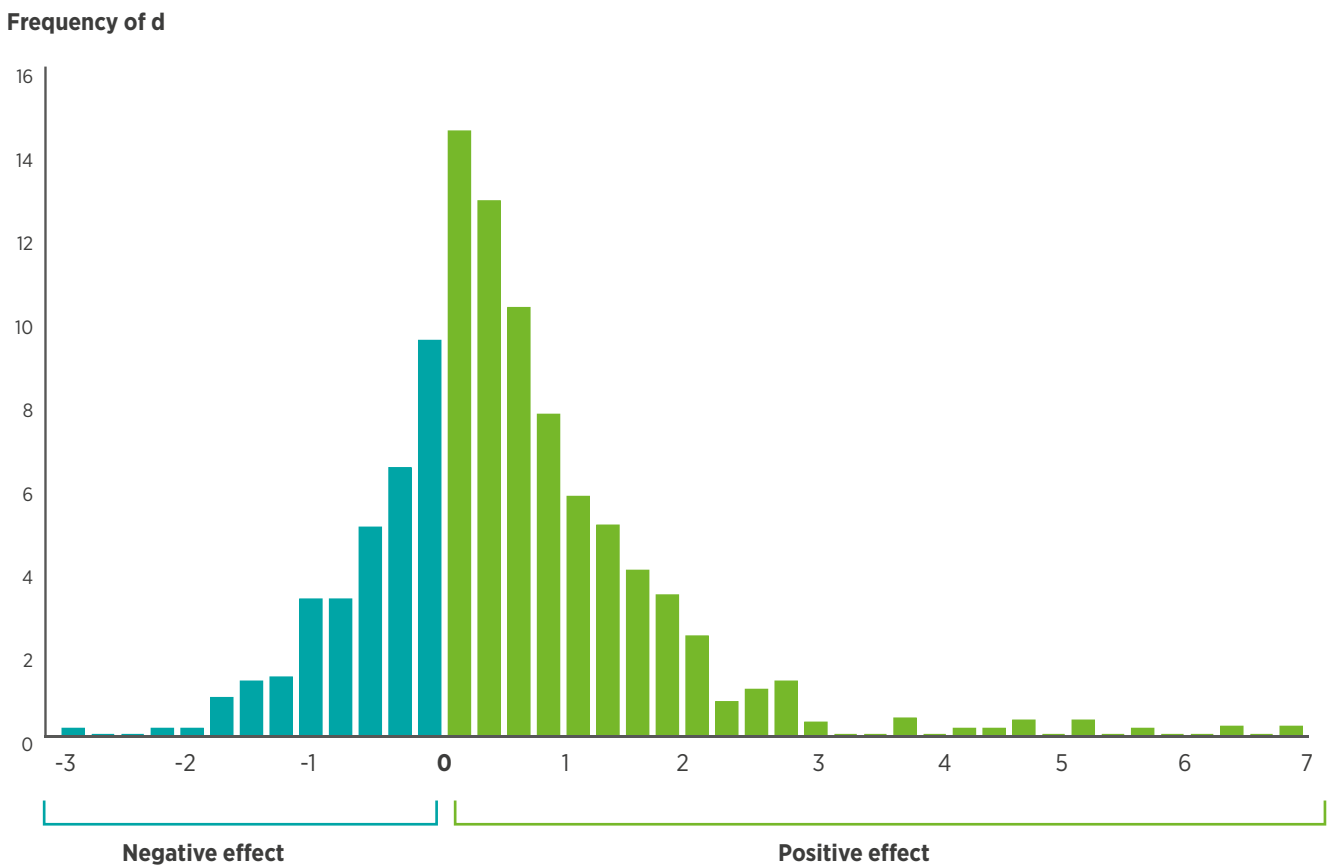
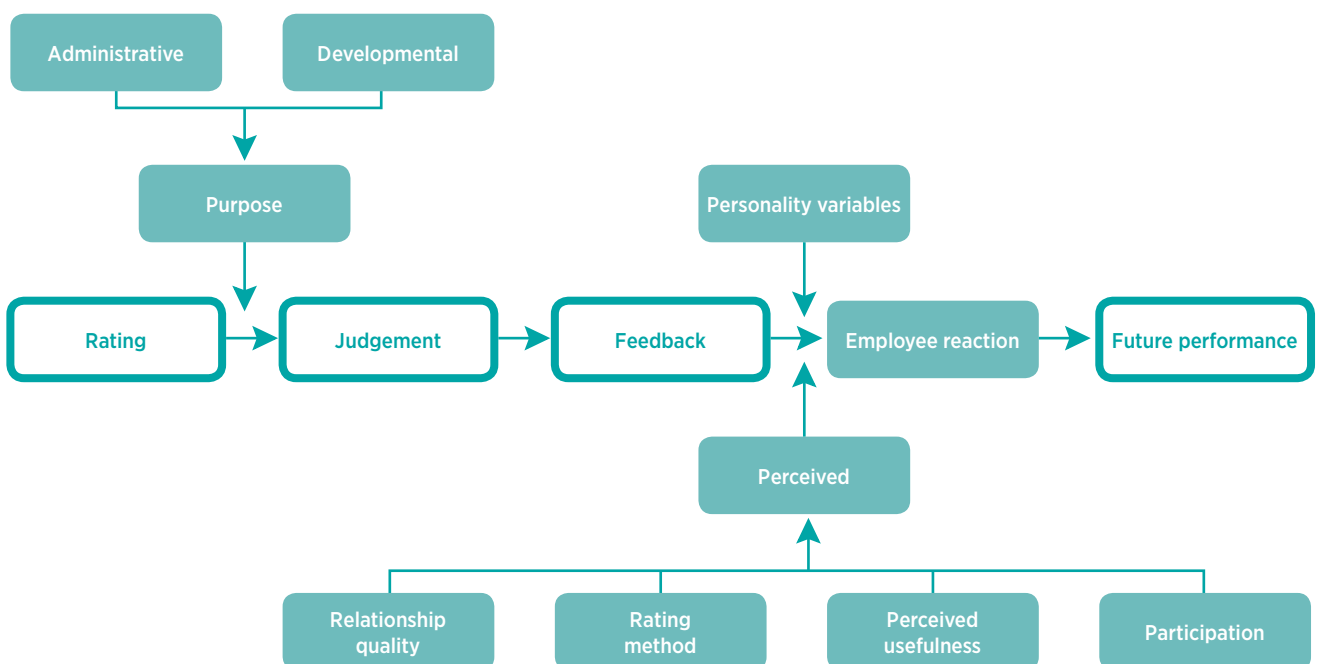


Figure 3: Moderators and mediators



Question 4: What is known about possible moderators and/or mediators that affect the relationship between performance appraisal and workplace performance?⁸

As previously stated, one of the primary purposes of performance appraisals is to provide employees with clear feedback, which is intended to positively affect workplace performance. However, in their meta-analysis, Kluger and DeNisi (1996) found that although performance feedback generally improves performance, in more than one third of studies, feedback actually lowered performance. Similar results have been reported in meta-analyses of multi-source feedback: some of the studies included reported performance improvements, while some did not, and others reported inconclusive results (Seifert et al 2003, Smither et al 2005). These findings suggest that the effect of performance appraisal is moderated and/or mediated by several factors (see Figure 3).

1 Reactions to feedback, rather than the feedback itself, influence performance (level A)

As previously stated, research has found that although feedback generally improves performance, in more than one third of the studies, feedback lowered performance. Several theoretical models propose that employees' reactions to feedback likely determine the extent to which they will use it to improve performance (for example, Ilgen et al 1979, Murphy and Cleveland 1995). Employees

have several behavioural options when confronted with a discrepancy between what they wish to achieve and the appraisal feedback received. For example, they can accept the feedback and put in more effort to improve their performance, but they can also reject the feedback, feel angry and/or disappointed, and shift their attention away from their tasks. In the meta-analysis by Kluger and DeNisi (1996), it was found that the last option is likely when the feedback threatens an employee's self-esteem. A similar finding is found in the meta-analysis by Smither et al (2005): employees who express positive emotions immediately after receiving feedback show higher performance ratings, but those who express negative emotions show lower performance ratings.

2 Personality variables moderate reaction to the feedback (level n/a)⁹

There is no doubt that personality variables moderate the reaction to (negative) feedback, but they fall outside the focus of this REA. Among the personality variables that are known to be involved in the reaction to feedback are self-esteem (for example, Ilgen et al 1979), locus of control (for example, Ilgen et al 1979), tendency for cognitive interference (Kuhl 1992, Mikulincer 1989), altruism (Korsgaard et al 1994) and openness to feedback (Smither et al 2005).

3 The perceived fairness of the performance appraisal process has a medium to large moderating effect on future performance (level A)

A fair process is widely regarded as a prerequisite for the effectiveness of performance appraisal, a construct that in academia is often referred to as procedural justice. This reflects *'the perceived fairness of decision-making processes and the degree to which they are consistent, accurate, unbiased, and open to voice and input'* (Colquitt et al 2013). Empirical research has demonstrated that when procedures are perceived as fair, reactions are favourable, largely irrespective of the outcome. This interaction effect is called the fair process effect and has been shown empirically in several studies in different contexts (for a review, see Brockner and Wiesenfeld 1996).

Surprisingly this REA yielded only two studies that directly examined the relationship between the perceived fairness of the performance appraisal procedure and future performance. A before-after study found that performance appraisal incorporating the principles of fairness and due process tends to positively affect employees' reactions to feedback and their resulting overall job performance (Jawahar 2010). In addition, a recent randomised controlled study confirmed this finding and demonstrated that employees' perceptions of fairness had an effect on the relationship between feedback and overall task performance (Budworth et al 2015).

However, as mentioned above, several meta-analyses have demonstrated that perceived procedural justice more generally has a medium to large moderating effect on organisational outcomes, such as performance,¹⁰ productivity, satisfaction and commitment (Cohen-Charash and Spector 2001, Viswesvaran et al 2002).

4 Both rating format and rating method have small to large moderating effects on perceived fairness, self-efficacy and ability to improve (level A)

A randomised controlled study (Bartol et al 2001) found that rating segmentation (that is, the number of alternative appraisal categories available for rating employee performance) affects employees' perception of fairness. More specifically, moderate segmentation (five categories) resulted in higher self-efficacy regarding employees' ability to improve their performance and higher goals than a low segmentation (three categories). Another randomised controlled study demonstrated that a substantially lower degree of fairness was reported when a forced distribution rating system was used for administrative purposes, especially when there was reduced variability in ratees' (actual) task performance (Schleicher et al 2009).

5 Feedback perceived as useful improves perceptions of fairness (level B*)

A four-year longitudinal controlled study found that feedback interviews perceived as useful improve perceptions of fairness. In contrast, when such interviews are perceived to be unhelpful, the impact on justice perception is negative (Linna et al 2012).

6 Negative feedback adversely affects perceived fairness (level C*), whereas feedback that focuses only on positive aspects has a medium positive effect on both perceived fairness and overall job performance (level A*)

The outcome of a longitudinal study suggests that employees who receive negative performance-appraisal feedback report lower perceptions of fairness. This effect even persists six months after the performance appraisal (Lam et al 2002). In addition, a recent randomised controlled study demonstrates that employees who receive feedback that focuses only on positive aspects (such as the employee's strength and accomplishments)¹¹ perform significantly better on the job four months later than employees who receive a traditional performance appraisal interview (Budworth et al 2015).

7 Participation has a medium to large moderating effect on perceived fairness (level B)

A meta-analysis of 32 studies (Cawley et al 1998) suggests that participation in the performance-appraisal process has a large, positive effect on perceived fairness of the appraisal, perceived utility and motivation to improve after the appraisal. This effect was most strongly for value-expressive participation (that is, for the sake of having one's 'voice' heard) than for instrumental participation (that is, for the purpose of influencing the end result). This finding is consistent with a large meta-analysis of more than 200 studies, demonstrating that voice has a medium positive effect on an employee's perception of procedural justice (Colquitt et al 2001)

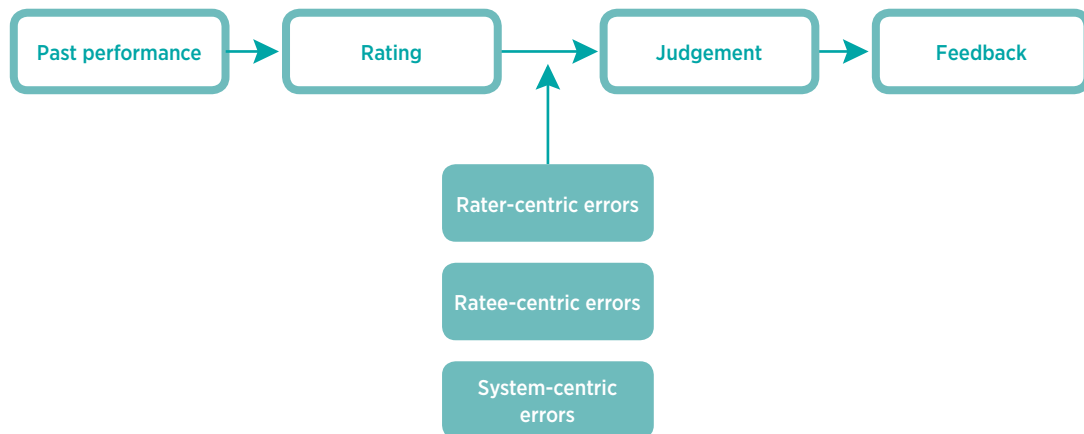
8 The quality of the relationship between manager and employee has a substantial moderating effect on the perceived fairness of the performance appraisal (level B)

A recent meta-analysis of 69 studies demonstrates that the quality of the relationship between the manager and the employee is strongly related to the employee's reaction to the performance appraisal (Pichler 2012). In addition, a longitudinal study finds that the quality of leader-member exchange (LMX) is a strong predictor for perceived fairness (Elicker et al 2006). LMX theory states that managers often have a special relationship with an inner circle of trusted employees, to whom they give higher levels of responsibility, decision influence and access to resources. In return, these employees work harder and are more committed to task objectives. The findings suggest that they are more likely to perceive the performance appraisal as fair.

Question 5: What is known about the reliability and validity of performance appraisal?

Performance appraisal is assumed to improve individual performance and organisational outcomes. To do this, it is essential that performance ratings are accurate and unbiased. However, there is a substantial body of research demonstrating that the accuracy of performance ratings can be influenced by a large number of different factors, bringing about poor rating quality and, as a result, affecting the reliability and validity of performance appraisal. The reliability and validity of performance ratings, also referred to as rating accuracy, have traditionally been assessed within the following three categories:

Figure 4: Rating accuracy



- 1 rater-centric rating errors (errors in judgement that occur when a person evaluates another person’s performance, for example rater bias and contrast effects)
- 2 ratee-centric rating errors (errors in judgement that occur because the person being evaluated deliberately influences the rater’s perception)
- 3 system-centric rating errors (errors in judgement that are due to flawed procedures or inaccurate rating scales).

Rater-centric rating errors

1 Employees’ contextual performance has a large positive effect on job performance ratings (level A)

Although task performance has been the traditional focus of research, individual work performance is considered to be more than meeting prescribed work goals. Researchers therefore distinguish an additional dimension of performance that is referred to as ‘contextual’ performance: extra-role behaviours in which employees go beyond their formal job requirements, such as taking on extra tasks, showing initiative or helping colleagues. Although several labels for this type of performance exist

(for example, organisational citizenship behaviour, extra-role performance or non-job-specific task performance), all refer to types of behaviour that go beyond the formally prescribed work goals (Koopmans et al 2011). Several controlled studies and meta-analyses have demonstrated that an employee’s contextual performance influences the perception of their overall job performance by their managers and, as a result, boosts performance ratings. Employees who voluntarily help others with work-related problems, make constructive suggestions to improve the efficiency of work processes, or co-operate with others to serve the interests of the organisation therefore tend to receive substantially higher performance ratings (Podsakoff et al 2013).

2 Managers’ implicit person theory regarding the malleability of personal attributes has a large effect on how they rate their employees (level A*)

Implicit person theory (IPT) concerns a person’s implicit beliefs about the malleability of personal attributes. Put differently, some people implicitly believe that personal attributes such as ability or behaviour are largely

fixed, whereas others think that personal attributes can change or develop over time. A randomised placebo-controlled study finds that managers who implicitly believe that personal attributes are fixed tend to give lower ratings for good performance when their employees have previously been given a negative performance rating (Heslin et al 2005). This finding suggests that such managers tend to pay less attention to the actual performance of an employee once they have formed an impression. Conversely, managers who implicitly believe that personal attributes can change or develop tend to give higher ratings for good performance when their employees have previously been given a negative performance rating, suggesting that they base the grading on a more conscientious consideration of the performance (instead of their initial impression). This finding explains why some managers acknowledge an improvement in an employee’s performance more than others.

3 Managers’ power level has a large to moderate effect on how they rate both others and themselves (level A)

A meta-analysis of 46 studies indicates that as a manager’s power level grows, their evaluation

of others becomes increasingly negative, whereas evaluations of themselves become ever more positive (Georges and Harris 1998). This finding suggests that performance evaluations by supervisors should be considered in light of their hierarchical position and power level.

4 Rater training has medium to large positive effects on rating accuracy (level A)

A meta-analysis of 29 studies demonstrates that rater training may have positive effects on rating accuracy (Woehr and Huffcutt 1994). Rater-error training and performance-dimension training both appear to be moderately effective at reducing halo error (the tendency to make inappropriate generalisations based on one aspect of a person's job performance) and somewhat less effective with respect to leniency (the tendency to evaluate all employees as outstanding and to give inflated ratings rather than true assessments of performance). Frame-of-reference training and behavioural-observation training appear to be the most effective (single) types of training, with results indicating a large positive effect on rating accuracy. A combination of different types of rater training showed mixed effects.

5 Male employees who experience a conflict between family and work receive lower performance ratings (level A*)

A randomised controlled study demonstrates that men who are involved in family caretaking events that result in an absence from the workplace receive lower overall performance ratings and lower reward recommendations than men who do not, whereas ratings of women are unaffected (Butler and Skattebo 2004). The gender of the rater(s) does not moderate the sex bias.

6 The outcome of managers' own performance appraisal has a large effect on the way in which they evaluate their employees (level A)

A combination of studies (including a randomised controlled study) demonstrated that managers who receive positive feedback about their performance subsequently rate their employees significantly higher than managers who receive negative feedback regarding their own performance (Latham et al 2008). Surprisingly, this effect even occurred when managers knew their own evaluation was bogus.

7 Introverted employees evaluate their extroverted and disagreeable colleagues' performance substantially lower (level A)

A cross-sectional study suggests that introverted (but not extroverted) peers consistently evaluate extroverted and disagreeable (but not introverted and agreeable) colleagues' performance as lower (Erez et al 2015). This finding is replicated by the same researchers in a randomised controlled study, which in addition demonstrated that introverts' sensitivity to the personal traits of other people and general impressions mediate this effect. This finding suggests that employees high in extroversion and disagreeableness¹² should be made aware that their trait-relevant behaviour may have a profoundly negative impact on how they are perceived by their introverted colleagues, which as a result may lead to reduced performance evaluation for collective accomplishments.

8 Whether or not an employee was hired or recommended by the rater has a large effect on performance rating (level A*)

Results from a randomised controlled study show that ratings are upwardly biased when participants rate the performance

of a person they had originally selected (Slaughter and Greguras 2008). This finding suggests that supervisors who are responsible for hiring employees rate them more favourably than (a) candidates whose pre-selection information they never viewed and (b) candidates they did not recommend for hiring. This bias in evaluations may unfairly reward or promote some employees, in that such rewards are not based on actual performance.

9 Rater liking has small to moderate effects on performance rating (level B)

A recent meta-analysis of 40 studies demonstrates that raters tend to evaluate those they like substantially more positively than those they dislike (Sutton et al 2013). The relationship between likeability and performance ratings, however, was weaker for ratings of organisational citizenship behaviour (OCB) than for ratings of task performance. In addition, the degree of likeability was more strongly related to supervisor and subordinate ratings than to peer ratings. Surprisingly, the relationship between likeability and performance ratings was not moderated by the purpose (for example administrative vs developmental) of the rating. These findings suggest that rater liking, for good or ill, appears to play a key role in the performance-rating process.

10 A rater's personality¹³ has small to medium effects on performance rating (level A/B)

A recent meta-analysis indicates that personality traits such as agreeableness, extroversion and emotional stability have a small to moderate positive effect on performance ratings (Harari et al 2015). In addition, a randomised controlled study shows that highly agreeable individuals tend to be

more lenient when rating people with poor performance, especially when they anticipate feedback/future collaboration (Randall and Sharples 2012). A controlled study, however, indicates that higher levels of conscientiousness are associated with lower performance ratings (Spence and Keeping 2010). In the meta-analysis the effect of rater personality is moderated by both purpose (attenuated when ratings are collected for administrative purposes and strengthened when ratings are collected for developmental purposes) and accountability (attenuated when accountability is high and strengthened when accountability is low). Cumulatively, raters' personality traits account for between 6% and 22% of the variance in performance ratings.

11 Gender bias has small to moderate effects on performance rating (Level B)

A meta-analysis of 32 studies shows little evidence of overall gender bias in performance appraisals in actual work settings (Bowen et al 2000). However, when only men served as raters, there were substantial pro-male biases. When the raters were a mix of men and women, the latter rated slightly more highly. In addition, measures considered masculine (for example leadership, implementation) produced a pro-male bias, while measures viewed as feminine (for example communication, interpersonal sensitivity) produced a pro-female bias.

12 Race bias has a small effect on performance rating (Level C)

A meta-analysis of 74 studies indicates that white raters assigned higher ratings to white ratees than to black ratees. Black raters also assigned higher ratings to black ratees than to white ratees (Kraiger and Ford 1985). This race

bias tended to be more likely in organisational settings where black employees composed a small percentage of the workforce.

13 There is a small positive effect of disability on performance rating for employees with a disability (Level A+)

A meta-analysis of 13 controlled studies shows an overall (small) positive effect of disability on performance evaluations of people with disabilities (Ren et al 2008). However, an overall (small) negative effect was found on both performance expectations and hiring decisions for people with disabilities.

Ratee-centric rating errors

1 Employees' tactics for influencing raters such as ingratiation or self-promotion have a moderate effect on performance ratings (Level C)

Several studies find that employees deliberately or unconsciously try to influence their manager to achieve a higher performance rating, especially when they face job insecurity (Huang et al 2013). Whereas theory has specified a number of types of tactics designed to influence, most of the research (Gordon 1996, Higgins et al 2003) tends to focus on the tactics of ingratiation (flattery and carrying out favours in order to enhance managerial liking) and self-promotion (appearing competent on the job and making managers aware of one's performance). A longitudinal study, however, indicates that self-promotion tactics have a negative effect on a manager's liking of the employee, which tends to result in a lower performance rating, whereas ingratiation tactics have a positive effect on a manager's liking of an employee, and subsequently result in a higher performance rating (Dulebohn et al 2004).

2 Employees' organisational citizenship behaviour has a moderate to large positive effect on performance ratings (Level A)

A recent meta-analysis based on 81 studies and a combined sample size of more than 31,000 employees indicates that organisational citizenship behaviour (OCB) has a medium to large effect on managers' evaluation of job performance (Podsakoff et al 2013). This effect is assumed to occur because: (a) managers may provide higher evaluations to employees exhibiting OCBs as a form of reciprocity (Podsakoff et al 1993); (b) OCBs are interpreted as behavioural manifestations of commitment and/or loyalty (Allen and Rush 2001); and (c) managers tend to like these individuals more (Lefkowitz 2000). Although this effect can be regarded as a form of rater bias, it is categorised as a ratee-centric effect, as OCB is sometimes used by employees to positively influence the outcome of the performance appraisal (Dulebohn et al 2005).

3 Employees' political skills have a small positive effect on performance ratings (Level C*)

A recent longitudinal study suggests that an employee's ability to effectively understand others at work, and use such knowledge to influence others to act in ways that enhance their personal and/or organisational objectives, is positively related to performance rating (Hung et al 2012). This ability, also referred to as political skills, was even found to moderate negative effects of employee behaviour. For instance, it was found that employees' voice behaviour (proactively challenging the status quo and making constructive suggestions) may have a negative effect on performance ratings. This negative effect, however,

was found to be moderated by an employee's political skills, suggesting that employees who have low levels of political skill are likely to experience the negative effects of voice behaviour on their performance ratings.

System-centric rating errors

1 The purpose of the performance appraisal moderates the performance rating (level A)

More than 60 years ago, Taylor and Wherry (1951) hypothesised that performance-appraisal ratings for administrative purposes, such as pay rises or promotions, would be more lenient than ratings obtained for employee-development purposes. Over the past decades this hypothesis has been confirmed in a large number of studies. A meta-analysis of 22 studies demonstrates that performance-appraisal ratings obtained for administrative purposes are, on average, one third of a standard deviation larger than those obtained for employee-development purposes (Jawahar and Williams 1997). In fact, randomised controlled studies have demonstrated that managers tend to apply a different decision process when making performance-evaluation decisions for administrative reasons (Pesta et al 2005). When presented with a performance-appraisal judgement for developmental reasons, managers are more likely to consider all the performance-related behaviours of the employee, and will use examples of both bad performance and good performance to reach their conclusions. When making promotion-related decisions, however, managers tend to consider only examples of poor performance and use a threshold to make the yes/no decision. As a result, employees may perceive the evaluative decision as less accurate and thus as unfair.

2 The reliability of individual performance measures depends on the type of measurement, the source of measurement and the level of job complexity (level A/B)

Meta-analyses have demonstrated that, in general, subjective and objective measures of employee performance are not interchangeable (Bommer et al 1995). For instance, in cases of complex jobs, objective measures lack test-retest reliability (Sturman et al 2005). In addition, it was found that employees' self-ratings tend to be higher than the rating of managers and peers (Harris and Schaubroeck 1988, Heidemeijer and Moser 2009). This finding confirms the outcome of previous research, indicating that employees tend to overestimate their own level of performance relative to that of others in the organisation (Harris and Schaubroeck 1988, Mabe and West 1982). Finally, several meta-analyses demonstrate that the inter-rater reliability of (subjective) performance ratings by peers tend to be lower than the reliability of ratings by managers, and that these reliabilities are even lower for complex jobs (Conway and Huffcutt 1997, Heidemeijer and Moser 2009) or for job dimensions that are difficult to measure (for example leadership, interpersonal competence) (Viswesvaran et al 1996, 2002).

3 Both rating format and rating method have small to large effects on the rating (level A/B)

A meta-analysis of 23 studies demonstrates that the correlation between (subjective) ratings of managers and (objective) performance measures is higher when a relative rating format is used (comparing the employee with other employees) than when an absolute rating format is used (comparing the employee with a standard). In addition, a higher correlation is found when a

composite rating method is used (a rating made on a multi-item scale with scores averaged to a final grade) than when an overall rating method is used (a rating made on a one-item scale) (Heneman 1986). Finally, a recent randomised controlled study demonstrates that performance ratings based on the consensus of multiple raters tend to be more accurate than individual ratings (Picardi 2015).

4 The medium used to report the outcome of the performance appraisal has a large effect on the rating (level A*)

A randomised controlled study demonstrates that raters tend to give more negative appraisals when using email than when using traditional paper-form methods (Kurtzberg et al 2005). In addition, a controlled study shows that employees evaluated with traditional paper-form methods report higher levels of quality for the ratings than employees evaluated with an online system (Payne et al 2009).

5 Accountability substantially affects both rating outcomes and rating accuracy (level A)

Several randomised controlled studies demonstrate that rating outcomes are affected when raters are (or feel) accountable for their rating. For instance, raters whose rating is to be checked by an expert provide substantially lower ratings relative to control raters (Roch 2005, Roch and McNall 2007). However, the opposite effect is found when managers have to justify their rating in a face-to-face meeting with the employee – these managers rate their employees substantially more positively (Klimoski and Inks 1990, Spence and Keeping 2010). Finally, it has been found (Palmer and Feldman 2005) that (perceived) accountability has positive effects on rating accuracy, such as contrast

effects (the tendency for a rater to evaluate a person relative to other individuals rather than on-the-job requirements, see also below) and halo effects (the tendency to make inappropriate generalisations from one aspect of a person's job performance).

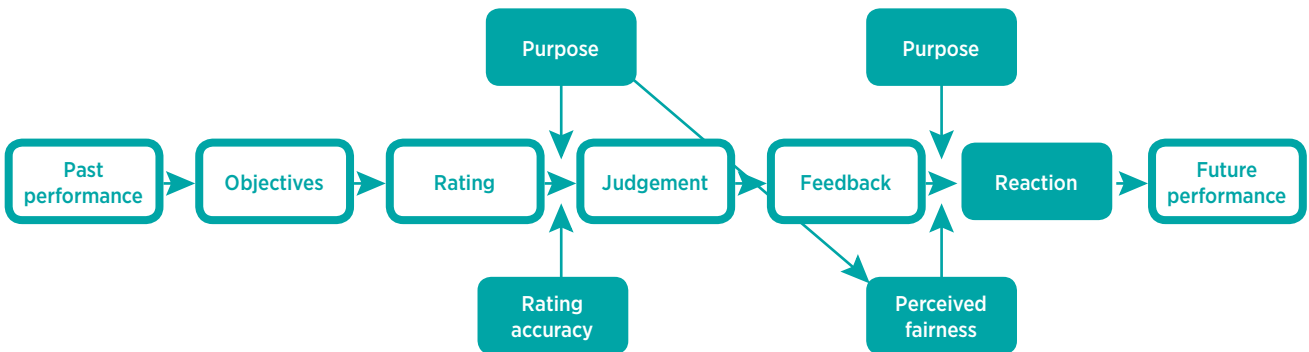
6 Contrast effects affect rating accuracy, but not always in a negative way (level A)

A contrast effect occurs when the performance of one employee (the anchor) has an effect on the evaluation of a subsequent performance of that employee or another employee (the target). For example, if the anchor performance is poor and the target performance is average, the contrast effect drives the evaluation

of the target performance in a positive direction. The finding of contrast effects is particularly robust in situations in which the anchor performance is either very poor or very good and the target performance is average (Smither et al 1988). Some researchers claim that contrast effects are an important source of rating error (Rowe 1967), whereas others maintain that the implications of contrast effects for actual rating situations are negligible (Hakel et al 1970). In fact, a randomised controlled study showed that the relationship between contrast effects and rating accuracy can even be positive,¹⁴ suggesting that not all rater-centric errors are an indicator of the quality of the rating (Becker and Miller 2002).

3 Synthesis

Figure 5: Synthesis of findings into the performance appraisal chain



‘... employee reaction to feedback is one of the most important mediators in the performance appraisal process’

Over the last 50 years, there has been a considerable number of studies on the topic of performance appraisal. In fact, performance appraisal may well be one of the most widely studied topics in the domain of management, with research on it dating back at least as far as the early 1920s, and continuing to the present day. After critically selecting and assessing the available empirical studies, we can conclude that the scientific evidence is rich in both quantity and quality.

In addition, the assumed positive effect of performance appraisal on work performance is grounded in three well-established social theories. However, this REA did not yield any randomised controlled studies that measured the direct effect of performance appraisal on workplace performance. Instead, the best available evidence consists of a large number of high-quality studies that focus on (one or multiple) separate elements of the appraisal process, such as rating, judgement, feedback or perceived fairness. The outcome of these studies is unequivocal

and suggests that the relationship between performance appraisal and work performance is contingent upon a large number of moderators and mediators.

The outcome of this REA indicates that employee reaction to feedback is one of the most important mediators in the performance-appraisal process. In fact, there is strong evidence that employees’ reaction to feedback, and not feedback *per se*, determines the extent to which their performance will improve. How an employee will react to the feedback on their appraisal, however, is strongly moderated by the perceived fairness of the appraisal process: when the procedure is perceived to be just, employee reactions are more likely to be favourable, largely irrespective of the outcome. Perceived fairness, in turn, is moderated by several other variables, of which perceived usefulness, rating method, rating accuracy, focus of the feedback, level of employee participation, and quality of the relationship between manager and employee seem to have the largest impact. A second

factor that moderates employees' reaction to the feedback embraces personality variables such as self-esteem, locus of control and openness to feedback.

In addition to these mediating and moderating factors, the outcome of this REA indicates that there is a wide number of variables that potentially affect the accuracy of the performance rating, which as a result may threaten the validity of the appraisal outcome (judgement) and seriously affect the fairness of the appraisal process. There is strong evidence that when a person evaluates another person's performance, systematic errors in judgement occur. Raters are often biased by a wide range of variables, such as implicit person theory, power level, relationship with the ratee, personality traits,

gender, and the outcome of their own performance appraisal. Raters, however, cannot be held responsible for all these errors. The evidence clearly indicates that ratees actively attempt to change a rater's judgement by using influence tactics such as ingratiation, self-promotion and political skills. In addition, errors in judgement occur because of the way in which the performance appraisal is conducted (procedures) or the performance is evaluated (measurement). Variables such as type and source of measurement, rating method, as well as aspects such as accountability and job complexity, may all affect the accuracy of the performance rating and, as a result, affect the validity and fairness of the appraisal process.

Finally, there is strong evidence that the purpose of the performance appraisal (administrative versus developmental) strongly affects the way in which raters evaluate and judge a person's performance. In fact, the evidence indicates that when making a performance-appraisal judgement for developmental reasons, raters apply a different decision-making process from when they are making a judgement for administrative reasons. As a result, employees may perceive the evaluative decision as less accurate and thus unfair.

Conclusion

Based on the evidence found, we conclude that performance appraisals can have positive effects on work performance, but that these effects are highly contingent upon a wide range of moderating factors. These include:

- 1 Reactions to feedback, rather than the feedback itself, influence performance (level A).
- 2 Personality variables moderate reaction to the feedback (level n/a).¹⁵
- 3 The perceived fairness of the performance appraisal process moderates the impact on future performance (medium to large effect; level A).
- 4 Both rating format and rating method moderate the effect on perceived fairness, self-efficacy and ability to improve (small to large effect; level A).
- 5 Feedback perceived as useful improves perceptions of fairness (level B*).
- 6 Negative feedback adversely affects perceived fairness (level C*), whereas feedback that focuses only on positive aspects contributes to perceived fairness and overall job performance (medium effect; level A*).
- 7 Participation contributes to perceived fairness (medium to large effect; level B).
- 8 The quality of the relationship between manager and employee contributes to the perceived fairness of the appraisal (substantial effect; level B).

We also identify a range of factors that can undermine or strengthen the reliability of performance

measures in appraisals. These are grouped into three categories. First, rater-centric errors include:

- 1 Employees' contextual performance influences their performance ratings (large effect; level A).
- 2 Managers' implicit person theory regarding the malleability of personal attributes influences how they rate their employees (large effect; level A*).
- 3 Managers' power level influences how they rate both others and themselves (large to moderate effect; level A).
- 4 Rater training contributes to rating accuracy (medium to large effect; level A).
- 5 Male employees who experience a conflict between family and work receive lower performance ratings (level A*).
- 6 The outcome of managers' own performance appraisal influences how they evaluate their employees (large effect; level A).
- 7 Introverted employees evaluate their extroverted and disagreeable colleagues' performance lower (large effect; level A).
- 8 Whether or not an employee was hired or recommended by the rater influences the performance rating (large effect; level A*).
- 9 If raters like people, they rate their performance higher (small to moderate effect; level B).
- 10 A rater's personality¹⁶ influences performance ratings (small to medium effect; level A/B).
- 11 Gender bias influences performance ratings (small to moderate effect; level B).

- 12 Race bias influences performance ratings (small effect; level C).
- 13 The performance of employees with a disability is rated higher (small effect; level A+).

Ratee-centric rating errors include:

- 1 Employees' tactics for influencing raters, such as ingratiation or self-promotion, affect performance ratings (moderate effect; level C).
- 2 Employees' organisational citizenship behaviour contributes to performance ratings (moderate to large effect; level A).
- 3 Employees' political skills contribute to performance ratings (small effect; level C*).

Finally, system-centric rating errors include:

- 1 The purpose of the appraisal moderates the performance rating (level A).
- 2 The type of measurement, source of measurement and level of job complexity affect the reliability of performance measures (level A/B).
- 3 Both rating format and rating method influence how positive ratings are (small to large effect; level A/B).
- 4 The medium used to report the outcome of the appraisal influences the performance rating (large effect; level A*).
- 5 Accountability affects both rating outcomes and rating accuracy (large effect; level A).
- 6 Contrast effects affect rating accuracy, but not always in a

Limitations

negative way (level A). This REA aims to provide a balanced assessment of what is known in the scientific literature about the effects of performance appraisal on individual work performance by using the systematic review method to search and critically appraise empirical studies. However, in order to be 'rapid', concessions were made in relation to the breadth and depth of the search process, such as the exclusion of unpublished studies, the use of a limited number of databases and a focus on empirical research published in the period 1980 to 2016 for meta-analyses and the period 2000 to 2016 for primary studies. In addition, the search for empirical studies was based only on terms such as 'performance appraisal', 'performance review', 'employee evaluation', and so on, and related terms such as 'feedback', 'judgement', or 'perceived fairness' were not included. As a consequence, some relevant studies may have been missed.

A second limitation concerns the critical appraisal of the studies included, which did not incorporate a comprehensive review of the psychometric properties of the tests, scales and questionnaires used. In addition, it should be noted that most of the studies included used performance ratings as an outcome measure, not actual performance, so the evidence is often indirect.

A third limitation concerns the fact that the evidence on several moderators is based on only one study (findings marked with an asterisk). Although most of these studies were well controlled or even randomised, no single study can be considered to be strong evidence – it is merely indicative.

Finally, this REA focused only on high-quality studies, that is, studies with a control group and/or a before-and-after measurement. For this reason, more than 50 cross-sectional studies were excluded. As a consequence, new, promising findings that are relevant for practice may have been missed.

Given these limitations, care must be taken not to present the findings presented in this REA as conclusive.

Endnotes

- 1 All reports are available at cipd.co.uk/coulddobetter
- 2 See www.cipd.co.uk/knowledge/journals
- 3 It should be noted that randomised controlled studies are often conducted in an artificial (lab-type) setting – with students carrying out prescribed work tasks – which may restrict their generalisability. Non-randomised studies in a field setting – with employees carrying out their normal tasks within an organisational setting – on the other hand, have a lower level of trustworthiness, but can still be useful for management practice.
- 4 In a meta-analysis, statistical analysis techniques are used to pool the results of individual studies numerically in order to achieve a more accurate estimate of the effect. Most studies defined as systematic reviews include a meta-analysis. The difference between a systematic review and a meta-analysis is therefore mainly semantic. Indeed, in medicine a meta-analysis is often called a systematic review.
- 5 It should be noticed that in feedback intervention theory, the definition of ‘performance’ includes a wide spectrum of tasks, such as physical tasks, cognitive tasks, complying with regulations and so forth.
- 6 It would, of course, make sense to differentiate between the performance of the individual, team and organisation. However, in most of the studies included, only individual performance is taken into account.
- 7 One may ask what type of performance these studies measured. After all, in performance-appraisal programmes, ratings can be collected on many performance dimensions (for example communication, risk-taking, customer focus, sales, decisiveness, and so on). Both meta-analyses included a large number of primary studies that examined the effect of feedback interventions on performance. In the meta-analysis by Smither et al, several studies reported separate scores for multiple dimensions, but these scores turned out to be highly correlated. In the meta-analyses by Kluger and DeNisi, a wide range of performance dimensions were taken into account (for example physical tasks, knowledge tasks, creativity, new tasks, rule-following, vigilance tasks, quality vs quantity, and so on), but only stronger effects were found for memory tasks and weaker effects for physical tasks and adherence to regulations. This suggests that the findings of these meta-analyses may apply to a wide range of performance outcomes.
- 8 A moderator is a variable that affects the direction and/or strength of the relation between an independent or predictor variable (in this case performance appraisal) and an outcome variable (work performance). Put differently, moderators indicate when or under what conditions a particular effect can be expected. For this reason, they are also referred to as ‘boundary conditions’. A mediator, by contrast, is a variable that specifies how or why a particular effect or relationship occurs. Thus, if you remove the effect of the mediator, the relationship between the independent or predictor variable (in this case performance appraisal) and the outcome variable (work performance) will no longer exist. In short, moderators specify when a certain effect will hold, whereas mediators determine how or why the effect occurs. The moderators and mediators are presented here in order of evidence quality and effect size, with the highest-quality evidence and greatest effect first.
- 9 The studies mentioned here are not included in this REA, so their quality was not evaluated.
- 10 Because of the large number of contexts in which perceived fairness was studied, the studies included in these meta-analyses used a wide variety of performance measures that came from various sources. For example, the measure of work performance included official performance ratings as they appeared in organisational files, and in-role behaviour ratings. A separate analysis was conducted for organisational citizenship behaviour.
- 11 This type of feedback is also known as ‘feedforward’ (see Kluger and Nir 2010).
- 12 Extroversion and agreeableness are two separate personality traits. Extroversion tends to be manifested in outgoing, talkative, energetic behaviour, whereas introversion is manifested in a more reserved and solitary approach. Agreeableness is a personality trait manifesting itself in individual behavioural characteristics that are perceived as kind, sympathetic, co-operative, warm and considerate.
- 13 One may wonder how personality can be measured. There is a wide consensus among researchers that the best way to describe personality is by using the ‘Big Five’ personality traits. These are: openness (inventive/curious vs consistent/cautious), conscientiousness (efficient/organised vs easy-going/careless), extroversion (outgoing/energetic vs solitary/reserved), agreeableness (friendly/compassionate vs analytical/detached) and neuroticism (sensitive/nervous vs secure/confident). The validity of the Big Five factors has been replicated numerous times in different languages and cultural contexts, and several validated measurement tools exist (for example the NEO-PI-R). Other measurements of personality, such as the Myers-Briggs Type Indicator (MBTI), are widely used, but their stability, validity and reliability are questionable at best.
- 14 Until now, no plausible explanation for this finding has been found.
- 15 The studies mentioned here are not included in this REA, so their quality was not evaluated.
- 16 See note 13 above.

References

- ADAMS, J.S. (1965) Inequity in social exchange. *Advances in Experimental Social Psychology*. Vol 2. pp267–99.
- ALLEN, T.D., and RUSH, M.C. (2001) The influence of ratee gender on ratings of organizational citizenship behavior. *Journal of Applied Social Psychology*. Vol 31, No 12. pp2561–87.
- BARENDT, E., JANSSEN, B. and VELGHE, C. (2016) *Rapid evidence assessment of the research literature on the effect of goal setting on workplace performance*. London: Chartered Institute of Personnel and Development. Available at: www.cipd.co.uk/coulddobetter. [Accessed 22 November 2016].
- BARTOL, K.M., DURHAM, C.C. and POON, J.M. (2001) Influence of performance evaluation rating segmentation on motivation and fairness perceptions. *Journal of Applied Psychology*. Vol 86, No 6. pp1106–19.
- BECKER, G. and MILLER, C. (2002) Examining contrast effects in performance appraisals: using appropriate controls and assessing accuracy. *Journal of Psychology*. Vol 136, No 6. pp667–83.
- BOMMER, W.H., JOHNSON, J.L., RICH, G.A., PODSAKOFF, P.M. and MACKENZIE, S.B. (1995) On the interchangeability of objective and subjective measures of employee performance: A metaanalysis. *Personnel Psychology*. Vol 48, No 3. pp587–605.
- BOWEN, C.-C., SWIM, J.K. and JACOBS, R.R. (2000) Evaluating gender biases on actual job performance of real people: a meta-analysis. *Journal of Applied Social Psychology*. Vol 30, No 10. pp2194–2215.
- BROCKNER, J. and WIESENFELD, B.M. (1996) An integrative framework for explaining reactions to decisions: interactive effects of outcomes and procedures. *Psychological Bulletin*. Vol 120, No 2. p189.
- BUDWORTH, M.H., LATHAM, G.P. and MANROOP, L. (2015) Looking forward to performance improvement: a field test of the feedforward interview for performance management. *Human Resource Management*. Vol 54, No 1. pp45–54.
- BUTLER, A.B. and SKATTEBO, A. (2004) What is acceptable for women may not be for men: the effect of family conflicts with work on job-performance ratings. *Journal of Occupational and Organizational Psychology*. Vol 77. pp553–64.
- CAWLEY, B.D., KEEPING, L.M. and LEVY, P.E. (1998) Participation in the performance appraisal process and employee reactions: a meta-analytic review of field investigations. *Journal of Applied Psychology*. Vol 83, No 3. pp615–33.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- COHEN-CHARASH, Y. and SPECTOR, P.E. (2001) The role of justice in organizations: a meta-analysis. *Organizational Behavior and Human Decision Processes*. Vol 86, No 2. pp278–321.
- COLQUITT, J.A., CONLON, D.E., WESSON, M.J., PORTER, C.O. and NG, K.Y. (2001) Justice at the millennium: a meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*. Vol 86, No 3. p425.
- COLQUITT, J.A., SCOTT, B.A., RODELL, J.B., LONG, D.M., ZAPATA, C.P., CONLON, D.E. and WESSON, M.J. (2013) Justice at the millennium, a decade later: a meta-analytic test of social exchange and affect-based perspectives. *Journal of Applied Psychology*. Vol 98, No 2. p199.
- CONWAY, J.M. and HUFFCUTT, A.I. (1997) Psychometric properties of multisource performance ratings: a meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*. Vol 10, No 4. pp331–60.
- DONNELLY, J. and TROCHIM, W. (2007) *The research methods knowledge base*. Mason, Ohio: Atomic Dog Publishing.
- DENISI, A.S. and PRITCHARD, R.D. (2006) Performance appraisal, performance management and improving individual performance: a motivational framework. *Management and Organization Review*. Vol 2, No 2. pp253–77.

- DULEBOHN, J.H., SHORE, L.M., KUNZE, M. and DOOKERAN, D. (2005) The differential impact of OCBS and influence tactics on leader reward behavior and performance ratings over time. *Organizational Analysis*. Vol 13, No 1. pp73–90.
- DULEBOHN, J.H., MURRAY, B. and FERRIS, G.R. (2004) The vicious and virtuous cycles of influence tactic use and performance evaluation outcomes. *Organizational Analysis*. Vol 12, No 1. pp53–74.
- ELICKER, J.D., LEVY, P.E. and HALL, R.J. (2006) The role of leader-member exchange in the performance appraisal process. *Journal of Management*. Vol 32, No 4. pp531–51.
- EREZ, A., SCHILPZAND, P., LEAVITT, K., WOOLUM, A.H. and JUDGE, T.A. (2015) Inherently relational: interactions between peers' and individual's personalities impact reward giving and appraisal of individual performance. *Academy of Management Journal*. Vol 58, No 6. p1761.
- FESTINGER, L. (1954) A theory of social comparison processes. *Human Relations*. Vol 7, No 2. pp117–140.
- FOLGER, R., KONOVSKY, M.A. and CROPANZANO, R. (1992) A due process metaphor for performance appraisal. *Research in Organizational Behavior*. Vol 14. p129.
- GEORGESEN, J.C. and HARRIS, M.J. (1998) Why's my boss always holding me down? A meta-analysis of power effects on performance evaluations. *Personality and Social Psychology Review*. Vol 2, No 3. pp184–95.
- GIFFORD, J. (2016) *In search of the best available evidence*. London: Chartered Institute of Personnel and Development. Available at: www.cipd.co.uk/knowledge/strategy/analytics [Accessed 23 November 2016].
- GORDON, R.A. (1996) Impact of ingratiation on judgments and evaluations: a meta-analytic investigation. *Journal of Personality and Social Psychology*. Vol 71. pp54–70.
- GRIFFIN, R. and EBERT, R.J. (2004) *Business essentials*. London: Prentice Hall.
- HAKEL, M.D., OHNESORGE, J.P., and DUNNETTE, M.D. (1970). Interviewer evaluations of job applicants' resumes as a function of the qualifications of the immediately preceding applicants: an examination of contrast effects. *Journal of Applied Psychology*, Vol 54, No1. p27.
- HALBE, S. (1951) Appraisal of job performance. *National Industrial Conference Board*. Vol 121.
- HARARI, M.B., RUDOLPH, C.W. and LAGINESS, A.J. (2015) Does rater personality matter? A meta analysis of rater Big Five performance rating relationships. *Journal of Occupational and Organizational Psychology*. Vol 88, No 2. pp387–414.
- HARRIS, M.M. and SCHAUBROECK, J. (1988) A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*. Vol 41, No 1. pp43–62.
- HEIDEMEIJER, H. and MOSER, K. (2009) Self-other agreement in job performance ratings: a meta-analytic test of a process model. *Journal of Applied Psychology*. Vol 94, No 2. pp353–70.
- HENEMAN, R.L. (1986) The relationship between supervisory ratings and results-oriented measures of performance: a meta-analysis. *Personnel Psychology*. Vol 39, No 4. p811.
- HESLIN, P.A., LATHAM, G.P. and VANDEWALLE, D. (2005) The effect of implicit person theory on performance appraisals. *Journal of Applied Psychology*. Vol 90, No 5. pp842–56.
- HIGGINS, C.A., JUDGE, T.A. and FERRIS, G.R. (2003) Influence tactics and work outcomes: a meta-analysis. *Journal of Organizational Behavior*. Vol 24. pp89–106.
- HOFFMAN, B., GORMAN, A., BLAIR, C., MERIAC, J., OVERSTREET, B. and ATCHLEY, K. (2012) Evidence for the effectiveness of an alternative multisource performance rating methodology. *Personnel Psychology*. Vol 65. pp531–63.
- HUANG, G.-H., XIONG-YING, N., ZHAO, H.H., ASHFORD, S.J. and LEE, C. (2013) Reducing job insecurity and increasing performance rating impression management matter? *Journal of Applied Psychology*. Vol 98, No 5. pp852–62.

- HUNG, H.-K., YEH, R.-S. and SHIH, H.-Y. (2012) Voice behavior and performance ratings: the role of political skill. *International Journal of Hospitality Management*. Vol 31, No 2. pp442–50.
- ILGEN, D.R., FISHER, C.D. and TAYLOR, M.S. (1979) Consequences of individual feedback on behavior in organization. *Journal of Applied Psychology*. Vol 64. pp349–71.
- ILGEN, D., BARNES-FARELL, J. and MCKELLIN, D. (1993) Performance appraisal process in the 1980s: what has it contributed to appraisals in use? *Organizational Behaviour and Human Decision Processes*. Vol 54. pp321–68.
- IQBAL, M.Z., AKBAR, S. and BUDHWAR, P. (2015) Effectiveness of performance appraisal: an integrated framework. *International Journal of Management Reviews*. Vol 17, No 4, pp510–33.
- JAWAHAR, I.M. (2010) The mediating role of appraisal feedback reactions on the relationship between rater feedback-related behaviors and ratee performance. *Group and Organization Management*. Vol 35, No 4. pp494–526.
- JAWAHAR, I.M. and WILLIAMS, C. (1997) Where all the children are above average: the performance appraisal purpose effect. *Personnel Psychology*. Vol 50.
- KLIMOSKI, R. and INKS, L. (1990) Accountability forces in performance appraisal. *Organizational Behavior and Human Decision Processes*. Vol 45. pp194–208.
- KLUGER, A.N. and DENISI, A. (1996) The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*. Vol 119, No 2. p254.
- KLUGER, A.N. and NIR, D. (2010) The feedforward interview. *Human Resource Management Review*. Vol 20. pp235–46.
- KOOPMANS, L., BERNAARDS, C.M., HILDEBRANDT, V.H., SCHAUFELI, W.B., DE VET HENRICA, C.W. and VAN DER BEEK, A.J. (2011) Conceptual frameworks of individual work performance: a systematic review. *Journal of Occupational and Environmental Medicine*. Vol 53, No 8. pp856–66.
- KORSGAARD, A., MEGLINO, B.M. and LESTER, S.W. (1994, August) *The virtue of being altruistic: the role of the value of helping and concern in individuals' reactions to feedback from others*. Paper presented at the 1994 meeting of the Academy of Management, Dallas, TX.
- KRAIGER, K. and FORD, J.K. (1985) A meta-analysis of ratee race effects in performance ratings. *Journal of Applied Psychology*. Vol 70, No 1. pp56–65.
- KUHL, J. (1992) A theory of self-regulation: actions vs. state orientation, self-discrimination, and some applications. *Applied Psychology: An International Review*. Vol 41. pp97–129.
- KURTZBERG, T.R., NAQUIN, C.E. and BELKIN, L.Y. (2005) Electronic performance appraisals: the effects of e-mail communication on peer ratings in actual and simulated environments. *Organizational Behavior and Human Decision Processes*. Vol 98, No 2. pp216–26.
- LAM, S., YIK, M. and SCHAUBROECK, J. (2002) Responses to formal performance appraisal feedback: the role of negative affectivity. *Journal of Applied Psychology*. Vol 87, No 1. pp192–201.
- LATHAM, G.P., BUDWORTH, M.-H., YANAR, B. and WHYTE, G. (2008) The influence of a manager's own performance appraisal on the evaluation of others. *International Journal of Selection and Assessment*. Vol 16, No 3. pp220–28.
- LEFKOWITZ, J. (2000) The role of interpersonal affective regard in supervisory performance ratings: a literature review and proposed causal model. *Journal of Occupational and Organizational Psychology*. Vol 73. pp67–85.
- LINNA, E.M., VAN DEN BOS, K., KIVIMAKI, M., PENTTI, J. and VAHTERA, J. (2012) Can usefulness of performance appraisal interviews change organizational justice perceptions? A 4-year longitudinal study among public sector employees. *International Journal of Human Resource Management*. Vol 23, No 7. pp1360–75.
- MABE, P. and WEST, S. (1982) Validity of self-evaluation of ability: a review and meta-analysis. *Journal of Applied Psychology*. Vol 67, No 3. pp280–96.

- MIKULINCER, M. (1989) Cognitive interference and learned helplessness: the effects of off-task cognitions on performance following unsolvable problems. *Journal of Personality and Social Psychology*. Vol 57. pp129–35.
- MURPHY, K.R. and CLEVELAND, J. (1995) *Understanding performance appraisal: social, organizational, and goal-based perspectives*. London: Sage.
- PALMER, J. and FELDMAN, J. (2005) Accountability and need for cognition effects on contrast, halo and accuracy in performance ratings. *Journal of Psychology*. Vol 136, No 2. pp119–37.
- PAYNE, S.C., HORNER, M.T., BOSWELL, W.R., SCHROEDER, A.N. and STINE-CHEYNE, K.J. (2009) Comparison of online and traditional performance appraisal systems. *Journal of Managerial Psychology*. Vol 24, No 6. pp526–44.
- PESTA, B., KASS, D. and DUNNEGAN, K. (2005) Image theory and the appraisal of employee performance: to screen or not to screen? *Journal of Business and Psychology*. Vol 19, No 3.
- PETTICREW, M. and ROBERTS, H. (2006) How to appraise the studies: an introduction to assessing study quality. In: PETTICREW, M. and ROBERTS, H. (eds) *Systematic reviews in the social sciences: a practical guide* (pp125–63). Hoboken, NJ: Wiley.
- PETTIJOHN, C.E., PETTIJOHN, L.S. and D'AMICO, M. (2001) Characteristics of performance appraisals and their impact on sales force satisfaction. *Human Resource Development Quarterly*. Vol 12. pp127–46.
- PETTIJOHN, C., PETTIJOHN, L.S., TAYLOR, A.J. and KEILLOR, B.D. (2001) Are performance appraisals a bureaucratic exercise or can they be used to enhance sales-force satisfaction and commitment? *Psychology and Marketing*. Vol 18, No 4. pp337–64.
- PICARDI, C. (2015) The effects of multi-rater consensus on performance rating accuracy. *Journal of Strategic Human Resource Management*. Vol 4, No 2.
- PICHLER, S. (2012) The social context of performance appraisal and appraisal reactions: a meta-analysis. *Human Resource Management*. Vol 51, No 5. pp709–32.
- PODSAKOFF, N.P., WHITING, S.W., WELSH, D.T. and MAI, K.M. (2013) Surveying for 'artifacts': the susceptibility of the OCB–performance evaluation relationship to common rater, item, and measurement context effects. *Journal of Applied Psychology*. Vol 98, No 5. pp863–74.
- PODSAKOFF, P.M., MACKENZIE, S.B. and HUI, C. (1993) Organizational citizenship behaviors and managerial evaluations of employee performance: a review and suggestions for future research. In: FERRIS, G.R. and ROWLAND, K.M. *Research in personnel and human resources management* (pp1–40). Greenwich, CT: JAI Press.
- RANDALL, R. and SHARPLES, D. (2012) The impact of rater agreeableness and rating context on the evaluation of poor performance. *Journal of Occupational and Organizational Psychology*. Vol 85, No 1. pp42–59.
- REN, L.R., PAETZOLD, R.L. and COLELLA, A. (2008) A meta-analysis of experimental studies on the effects of disability on human resource judgments. *Human Resource Management Review*. Vol 18, No 3. pp191–203.
- ROCH, S. (2005) An investigation of motivational factors influencing performance ratings. *Journal of Managerial Psychology*. Vol 20, No 8. pp695–711.
- ROCH, S. and MCNALL, L. (2007) An investigation of factors influencing accountability and performance ratings. *Journal of Psychology*. Vol 141, No 5. pp499–523.
- ROWE, P.M. (1967) Order effects in assessment decisions. *Journal of Applied Psychology*. Vol 51, No 2. p170.
- ROWLAND, C.A. and HALL, R.D. (2012) Organizational justice and performance: is appraisal fair? *EuroMed Journal of Business*. Vol 7, No 3. pp280–93.
- SCHLEICHER, D.J., BULL, R.A. and GREEN, S.G. (2009) Rater reactions to forced distribution rating systems. *Journal of Management*. Vol 35, No 4. pp899–927.
- SEIFERT, C.F., YUKL, G. and MCDONALD, R.A. (2003) Effects of multisource feedback and a feedback facilitator on the influence behavior of managers toward subordinates. *Journal of Applied Psychology*. Vol 88, No 3. p561.

- SHADISH, R., COOK, T.D. and CAMPBELL, D.T. (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton, Mifflin and Company.
- SHAUGHNESSY, J.J. and ZECHMEISTER, E.B. (1985) *Research methods in psychology*. New York: Alfred A. Knopf.
- SLAUGHTER, J.E. and GREGURAS, G.J. (2008) Bias in performance ratings: clarifying the role of positive versus negative escalation. *Human Performance*. Vol 21, No 4. pp414–26.
- SMITHER, J., LONDON, M. and REILLY, R. (2005) Does performance improve following multisource feedback? A theoretical model, meta-analysis and review of empirical findings. *Personnel Psychology*. Vol 58, No 1. pp33–66.
- SMITHER, J., REILLY, R. and BUDA, R. (1988) The effects of prior performance information on ratings of present performance: contrast versus assimilation revisited. *Journal of Applied Psychology*. Vol 73. pp487–96.
- SPENCE, J.R. and KEEPING, L.M. (2010) The impact of non-performance information on ratings of job performance: a policy-capturing approach. *Journal of Organizational Behavior*. Vol 31, No 4. p587.
- STURMAN, M.C., CHERAMIE, R.A. and CASHEN, L.H. (2005) The impact of job complexity and performance measurement on the temporal consistency, stability, and test-retest reliability of employee job performance ratings. *Journal of Applied Psychology*. Vol 90, No 2. pp269–83.
- SUTTON, A.W., BALDWIN, S.P., WOOD, L. and HOFFMAN, B.J. (2013) A meta-analysis of the relationship between rater liking and performance ratings. *Human Performance*. Vol 26, No 5. pp409–29.
- TAYLOR, E.K. and WHERRY, R.J. (1951) A study of leniency in two rating systems. *Personnel Psychology*. Vol 4, No 1. pp39–47.
- TJAHJONO, H.K. (2014) The fairness of organizations' performance appraisal social capital and the impact toward affective commitment. *International Journal of Administrative Science and Organization*. Vol 21, No 3.
- VISWESVARAN, C., ONES, D.S. and SCHMIDT, F.L. (1996) Comparative analysis of the reliability of performance ratings. *Journal of Applied Psychology*. Vol 81, No 5. pp557–74.
- VISWESVARAN, C., SCHMIDT, F.L. and ONES, D.S. (2002) The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance: unconfounding construct-level convergence and rating difficulty. *Journal of Applied Psychology*. Vol 87, No 2. pp345–54.
- VISWESVARAN, C., SCHMIDT, F.L. and ONES, D.S. (2005) Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*. Vol 90, No 1. pp108–31.
- WALSTER, E., WALSTER, G.W. and SCOTT, W.G. (1978) *Equity: theory and research*. Boston: Allyn and Bacon.
- WOEHR, D.J. and HUFFCUTT, A.I. (1994) Rater training for performance appraisal: a quantitative review. *Journal of Occupational and Organizational Psychology*. Vol 67, No 3. pp189–205.

Appendix 1

Search terms and hits

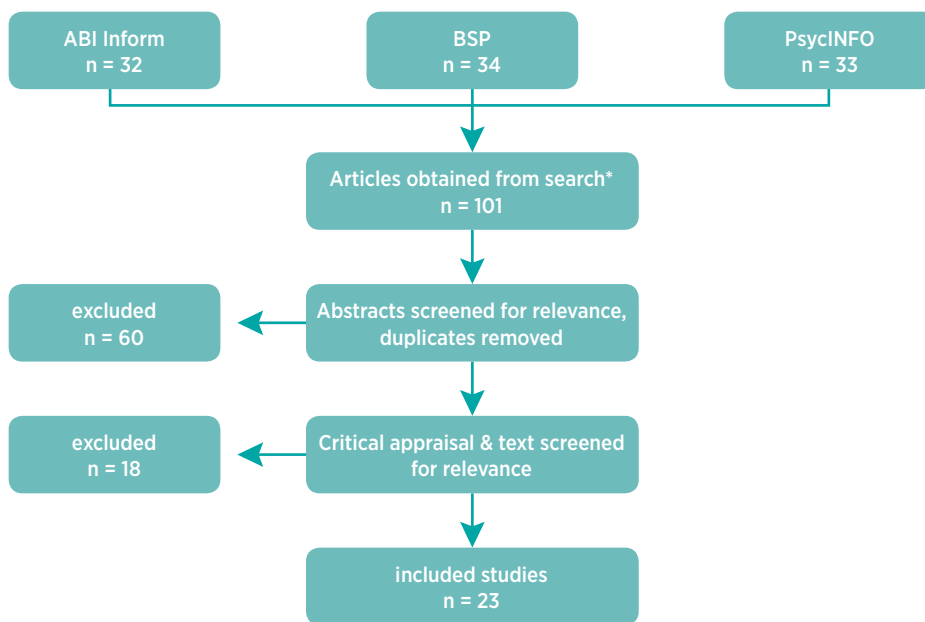
ABI/Inform Global, Business Source Elite, PsycINFO
Peer-reviewed, scholarly journals, May 2016

Search terms	ABI	BSP	PSY
S1: ti(performance) AND ti(apprais*)	706	883	577
S2: ab("performance appraisal*")	1,093	1,328	810
S - : ti(performance) AND ti(review*)	500+ > nr		
S3: ti("performance review*")	39	58	14
S4: ab("performance review")	150	289	50
S - : ti(performance) AND ti(evaluat*)	1500+ > nr		
S5: ti("performance evaluation*")	1,143	2,055	248
S6: ab("performance evaluation*")	2,464	4,274	727
S7: ti(performance) AND ti(rating*)	326	382	626
S8: ab(performance rating*)	160	217	197
S9: ti(employee) AND ti(apprais*)	81	100	61
S10: ab("employee appraisal")	16	22	21
S11: S1 - S10	4,936	7,887	2,489
S12: S11 AND filter ti(meta-analy*) OR ab(meta-analy*) OR ti("systematic review") OR ab("systematic review")	32	34	33
S13: S1 OR S3 (OR S5 - PsycINFO) OR S7 OR S9	1,060	1,318	1,432
S14: S13 AND filter ab(study OR studies OR empirical OR experiment* OR control* OR longitudinal) and limited to 2000 - 2016	244	299	-
S15: S13 AND filter quantitative study and limited to 2000 - 2016	-	-	444
S16: S14 AND filter ab(longitudinal) OR ab(experiment*) OR ab(control*)	40	52	62
S17: S14 OR S15 NOT S16	204	247	382

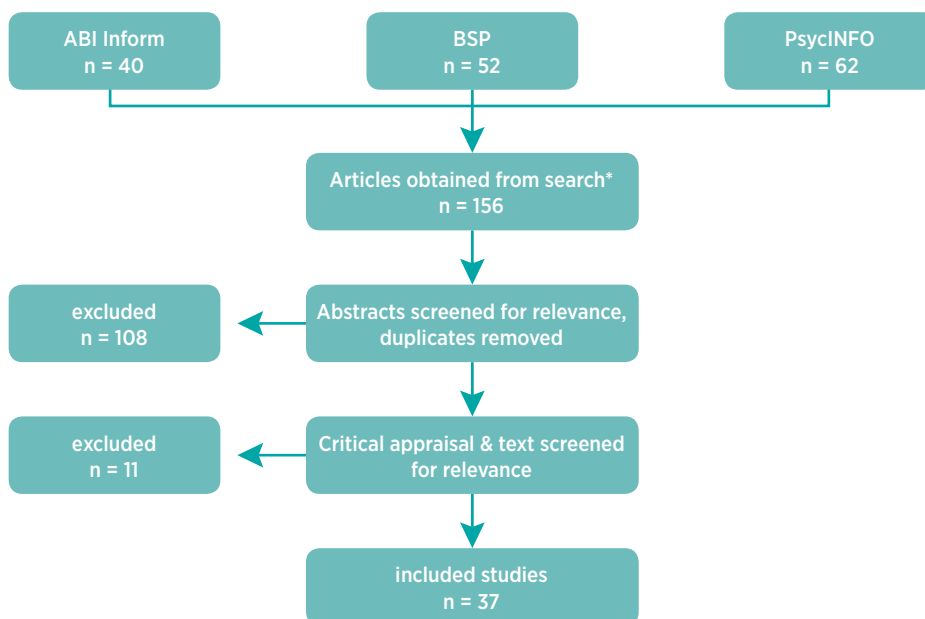
Appendix 2

Selection process

Meta-analyses or systematic reviews



Meta-analyses or systematic reviews



* The total number of articles obtained from the search include a few additional articles that were referenced in articles found (but not identified directly in our search) and judged to be worthy of inclusion.



CIPD

Chartered Institute of Personnel and Development
151 The Broadway London SW19 1JQ United Kingdom

T +44 (0)20 8612 6200 **F** +44 (0)20 8612 6201

E cipd@cipd.co.uk **W** cipd.co.uk

Incorporated by Royal Charter

Registered as a charity in England and Wales (1079797) and Scotland (SC045154)

Issued: December 2016 Reference: 7357 Appraisal © CIPD 2016