

Evidence Matters

The death of the
performance review:
fashion, fad or
forward-thinking?



Contents

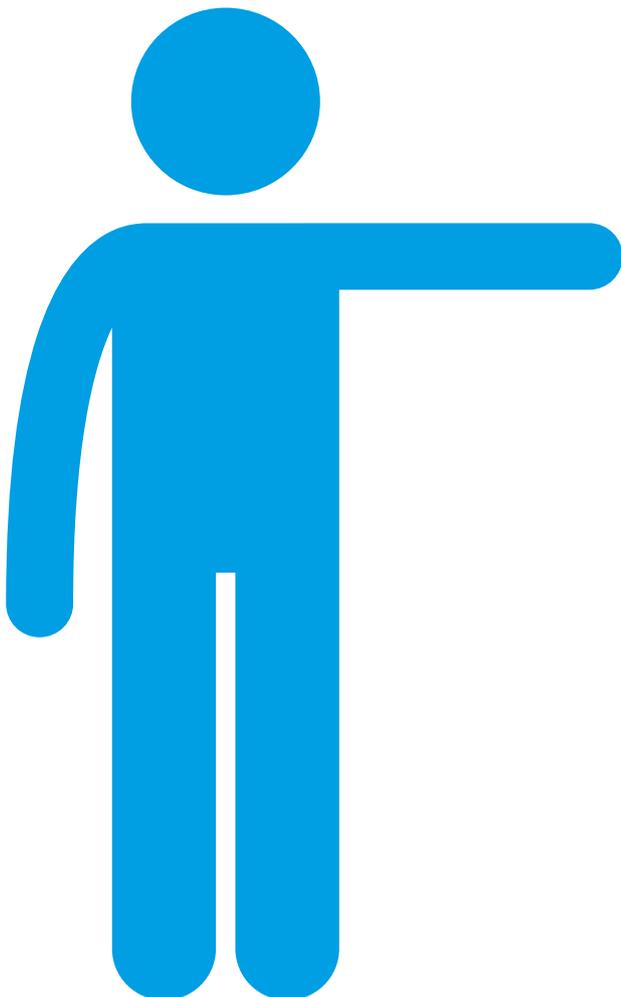
Introduction	3
Performance reviews: a systems perspective	4
What we did	5
The death of the performance review	6
Why do organisations conduct performance reviews?	7
Replacing performance reviews: some case studies	9
Adobe	10
Deloitte	11
Are we ‘throwing the baby out with the bath water’?	12
1. Goal setting	13
2. Feedback – how much and what type?	16
3. Rating performance	20
Returning to the systems perspective	23
Some key questions to consider	24
How can we help?	25
Understanding where you are now	25
Improving your current process	25
Designing and evaluating a new performance review process	25
References	26
About the Future Work Centre	28

We'd like to acknowledge the contribution of the following members of the Future Work Centre team in conducting this review and compiling this report: Claudia Nuttgens, Principal Psychologist, Dr Richard A MacKinnon, Insight Director, Claire Rahmatallah, Commercial Director, Rhiannon McGuinness, Insight Psychologist, Della Rath, Insight Psychologist.

Introduction

Recent coverage in the general and HR media has reported some high-profile decisions to abandon or replace the traditional performance review. At the Future Work Centre, our focus is on establishing the evidence supporting organisational decisions, so this news was of interest to us.

We wanted to understand whether these decisions were actually underpinned by evidence and data, or whether organisations might be engaging in an HR-related fad. Consider the potential negative outcomes of changing how employees' performance – and therefore reward – is assessed. We therefore thought it prudent to examine the evidence for the performance review and share our findings to facilitate evidence-based decisions.



Performance reviews: a systems perspective

If we view a typical organisation as being more like a complex ecosystem, rather than silos of discrete teams and functions, it soon becomes apparent that changes made to one element of the system can have unintended, knock-on impacts elsewhere. Just like dropping a pebble into a pond, its impact causes ripples that spread out from the source. Some disturbance to the water is visible, most isn't.

This metaphor is particularly relevant when we turn our attention to how employee performance is rated and how this links to motivation, satisfaction, loyalty and subsequent performance. Changes to performance management processes should be carried out with intention (e.g. "We want to improve the efficiency of the process" or "We want to gather more actionable data"), as well as being mindful of the possible negative consequences that such a change may cause.

It is entirely possible that a change to this part of the organisational system could be welcomed by employees, leading to increased satisfaction and, possibly, productivity. It is also possible that a change to how performance reviews are carried out could lead to increased suspicion of the entire performance management process, decreased satisfaction and eventually, increased staff turnover.

How can we determine what may happen as a result of our actions? One way is to consider intended and likely consequences at multiple levels:



- **Intra-personal:** What does this mean for individuals, considering the diversity of our employee population? Will any groups perceive this as a 'loss'? Will any groups be disadvantaged because of this change? In looking at the intra-personal level of the organisational system, we should consider factors such as gender, age, role, seniority and working arrangements. A change to performance management could impact line managers, if it requires them to adopt new behaviours or develop new performance management skills.
- **Inter-personal:** What might this change mean for relationships between employees? Relevant factors at this level of the system include peer relationships within a team, manager-subordinate relationships and how employees view the fairness of a system when they make comparisons between themselves and others.
- **Organisational:** What might this change mean for our organisation as a whole? The relevant factors here are the organisational history of effecting change to people-related processes, the organisation's culture and values. We should also consider the sector in which the organisation operates and how the rest of the sector carries out such activities.
- **Socio-economic:** Here we can consider factors like the national culture in which the organisation is operating, and whether changes to performance management can even be allowed by some cultures.

What we did

We undertook a rapid evidence review and an analysis of the trends in the uses of performance review. We looked for robust research and credible case studies that provided quality, reliable and generalisable evidence of what's really going on when organisations carry out performance reviews. We also examined what the HR press has been saying about this topic, to identify trends in the performance review 'narrative'.

This resulting analysis is unlikely to be perfect – there is no such thing as perfect evidence after all – but our review enables us to generate solid conclusions that should provide a springboard for informed discussions about what you do in your organisation.

The death of the performance review



Type this phrase into Google and your top hits consist of articles and papers published in 2015 by several well respected institutions and publications (Society for Human Resource Management, Financial Times, CIPD, Forbes, etc.). It's a trend that's hard to ignore if you're a practitioner, and your senior stakeholders may also be asking tricky questions about your current processes.

In 2014, CEB reported that 6% of Fortune 500 companies had either radically changed or dispensed with their performance review. And in a recent study, Deloitte found that 70% of their respondents were "currently evaluating" or have recently 'reviewed and updated' their performance management systems.

You might be considering joining them. Alternatively, you may feel satisfied with your current approach, but curious to know how it can be improved. Or, you may have run a process like this for some time, but have no idea what value it's adding to your organisation's performance.

Whatever your starting point, rather than doing as others have, or sticking to the status quo, it's worth thinking more deeply about performance review - what's working or what isn't, for whom, and in what situation.

Why do organisations conduct performance reviews?

The answer to this question may seem obvious, but our review of the research suggests that very few organisations are explicit about the purpose of the review process beyond its contribution to overall performance management. This is backed up in a paper by Jane Maley (2013) who found that a fundamental limitation of the performance review was its lack of clear purpose, and this lack of clarity was at all levels of the organisation, leading to confusion and conflict about how the resulting information is used.

Though it would seem sensible to assume that the ultimate aim of any people-focused process or strategy is to enhance overall business performance, the specific purposes of the performance review are varied. Organisations promote their approaches by emphasising a range of advantages and benefits including:

- Clarifying organisational goals, expectations, rules and standards.
- Allowing the manager to 'get to know' employees better.
- Creating an opportunity to provide feedback, identify what behaviours and skills an employee needs to develop, and identify any managerial action required.
- Providing a way of gathering information on performance that can be collated and used by the organisation.
- Collecting evidence to inform decisions on individual reward, role placement, etc.

These stated advantages are supported by logic, but rarely by evidence. We need to ask ourselves challenging questions about why we continue to use a process that can be expensive and is often loathed by managers and staff alike. These questions include:

- Is individual performance the right focus if we want to improve organisational performance? There is growing evidence that individual performance does not necessarily lead to organisational performance.
- How do you know that people are clearer about their goals because of the performance review?
- Are there better methods of ensuring that managers get to know their teams, and why is that important?
- Is the information you are collecting during the review accurate, fair and useful? And do you use that information properly?
- Does it matter that performance reviews are usually unpopular and often not conducted properly?
- Does getting feedback actually lead to better performance?

So why do we do performance reviews? Well, on the face of it, the answer is 'for lots of reasons'. These reasons are, at best, logical and aspirational and, at worse, probably not attainable or even linked to the performance review activity.

It's clear that many of the positively phrased justifications we come across are not supported by good evidence. This makes evaluation and evidence-based analysis of the performance review very difficult.

To help us answer the question, we needed to break down performance review into its key components and look at what works and what doesn't for each component.

Replacing performance reviews: some case studies

Before evaluating the evidence associated with performance review, we looked at some high profile examples of organisations that have replaced their annual performance review. We wanted to understand their rationales, explore their approaches and find out what evidence they had collected about the success, or otherwise, of the changes they had made.

Several large organisations such as Microsoft, Accenture and Gap, have dispensed with more traditional approaches to performance review and have published outcomes that make a compelling case for doing so. We summarise two of these case studies here.

An important consideration: taking an evidence-based perspective means that case studies should be interpreted and used with caution. When reading them, ask yourself the following questions.

- What measures are being used? Are those measures valid and reliable? For example, if an organisation reports an increase in satisfaction based on a measure in a staff survey, can we be sure that survey is actually measuring satisfaction in the first place?
- Why are there no case studies that report failures in new approaches to review and feedback? We may only see the successes in the public domain while the failures aren't explored.
- Who is sponsoring the study? Are the researchers impartial or is there a risk that they are focusing on evidence to prove a particular approach or increase positive perception of the brand?
- Be cautious about claims based on correlation. Correlation is not the same as causation. Could something other than the changes to the review system have caused the changes in attrition, the increase in satisfaction with feedback, and the other reported outcomes?
- Question even the most apparently logical inferences. For example, is it reasonable to conclude that reductions in attrition demonstrate that "employees felt valued, and that the check-ins had added elements of fairness and accuracy to Adobe's performance evaluation system"?



Adobe *(Adapted from a case study by Hinds, Sutton & Rao, Graduate School of Stanford Business, 2014, Case No. HR38)*

In 2012, Adobe's performance management system revolved around their annual reviews. It was estimated that during this time, 80,000 hours were spent on performance reviews. They have now moved from annual performance review to 'check-ins'. Their new approach encourages employees and managers to establish expectations and to share feedback on an ongoing basis.

Limitations of their annual reviews included:

- The 'stack ranking' approach appeared to lead to employee behaviours focused on ensuring that they were in the top 15%. This behaviour was viewed as sometimes being at the expense of the team.
- Adobe identified a recency effect, with managers focusing on what the employee had achieved in the past few months rather than summarising achievements evenly over the year. They also saw voluntary attrition rates escalate in February and March in the aftermath of the January reviews. Ratings below employee expectations were a primary stated cause of this attrition every year.
- Employees were far more focused on how they had been ranked than on the feedback they had received. They particularly dreaded the 'meets expectations' label.

Adobe replaced the system with a three-tiered 'check-in' framework involving:

- **Expectations:** Set, track and review clear expectations/goals and clarify roles, responsibilities and success criteria frequently throughout the year.
- **Feedback:** Give and receive ongoing feedback and coaching on a frequent and timely basis to recognise and address and/or improve performance against expectations.
- **Growth:** Provide opportunities to develop and increase skills, knowledge and experience in employee's current role in alignment with business needs and individual aspirations.

Results:

At the beginning of 2014, the system was deemed successful by Adobe. They reported noticeable changes at the employee, managerial and the overall organisational level.

Employee perspective: Because bad news was timely, it was not as catastrophic. In June 2014, two years after implementation, voluntary attrition continued to trend downward on a global basis. "This trend implied employees felt valued, and that the check-ins had added elements of fairness and accuracy to Adobe's performance evaluation system."

Manager perspective: Managers felt that there was less toll on them in terms of remembering information because they were consistently sharing information on performance.

Organisational perspective: Adobe's engagement survey showed strong results, with a significant 8% and 6% increase in the number of respondents agreeing with the statements "The feedback I receive is helpful to my performance" and "My manager is open to receiving feedback from me" respectively. However, the survey also showed that existing employees thought check-ins added, rather than decreased, stress because the bar was always set high and performance was constantly evaluated. Responses to bi-annual survey were promising: employees felt that the check-ins provided them with helpful insights into their performance and actionable feedback for improvement of performance. They reported that the check-in procedure fostered a culture of feedback, evidenced by the response rates for the survey which were 93% overall and up to 98% in some divisions.

Challenges of the new system:

- **Lack of training:** some employees did not fully understand the system, which led to differences between groups, in particular with regard to setting expectations.
- Some managers were still uncomfortable having difficult conversations with employees.
- Senior leaders were perceived as not being open to receiving feedback.

Key terms

Valid – do they measure what they say they measure?

Reliable – do they measure it consistently?



Deloitte (Adapted from article by Buckingham & Goodhall, Harvard Business Review, 2015)

Deloitte's new system has no cascading objectives, no annual reviews, and no 360 feedback tools. The hallmarks are speed, agility, one size fits one, constant learning, all underpinned by a new way of collecting reliable rich performance data. Deloitte state that

"Our question now is not 'What is the simplest view of you?' but 'What is the richest?'"

Why did Deloitte decide to make the change? They based their decision on three pieces of evidence:

1. A simple counting of hours – they estimated that 2 million hours were spent on the process.
2. A review of research in the science of ratings.
3. A 'carefully controlled' study of their own organisation.

Put simply, the process was too time consuming, expensive and though perceived as fair, end of year goals were seen as too restrictive.

Following a study of their high performing teams against a control group, Deloitte identified that these teams were strength-oriented – that is they tended to celebrate their successes and build on their strengths rather than focus on failures and issues associated with lack of competence. Therefore, their new system was designed to reflect this.

The objective of the new system is to fuel performance not manage it, and includes self-assessments of strengths and snapshots of performance throughout the year and at the end of projects. This information is then weighted according to the duration of a project and aggregated once a year.

And taking into account research evidence that suggests that people rate others inconsistently and

inaccurately, they also changed the emphasis of the measurement, to assess what team leaders would **do** with their team members not what they **think** of them. Team leaders are asked to rate employees against these four statements:

1. Given what I know of this person's performance, and if it were my money, I would award this person the highest possible compensation increase and bonus [measures overall performance and unique value to the organisation on a 5-point scale from "strongly agree" to "strongly disagree"].
2. Given what I know of this person's performance, I would always want him or her on my team [measures ability to work well with others on the same 5-point scale].
3. This person is at risk for low performance [identifies problems that might harm the customer or the team on a yes-or-no basis].
4. This person is ready for promotion today [measures potential on a yes-or-no basis].

As well as quarterly assessment, team leaders are required to hold weekly check-in meetings with each team member (initiated by the team member).

It's interesting to note that Deloitte chose these four questions and the project-based assessment, because that's what suits them and is meaningful to their staff. They do not explicitly advocate their approach for other organisations.

Results:

This approach is currently being piloted with 10% of employees. They are still experimenting with the levels of transparency throughout a year.

There is no doubt that the changes implemented by Adobe and Deloitte have been well thought through and have to some extent been based on internal organisational research, as well as the available evidence on some of the psychological and social processes involved.

But are a handful of high profile case studies enough to provide assurance that we should follow in their footsteps? How much can a small number of case studies tell us about the impact of performance reviews and how generalisable are their findings?

Are we ‘throwing the baby out with the bath water’?

This is a phrase used in an interesting opinion piece in the Sydney Morning Herald in 2013, by Peter Gahan (Director of the Centre for Workplace Leadership, University of Melbourne). He reflects on decisions to scrap performance reviews as a ‘management fad’ and cautions organisations to think carefully about whether it makes sense to scrap a process too hastily, when there is evidence that aspects of it can work.

We carried out our own review of the available scientific evidence in key practices that form part of traditional performance review. We focused our attention on three inter-related areas we view as being pivotal to the success or failure of performance review and appraisal:

- **Goal setting** – how we set specific goals and targets with the intention of motivating people to achieve the things the organisation needs. Most goal setting in organisations starts with the premise that stretching goals provoke higher levels of effort than an unfocussed ‘do your best’ approach.
- **Feedback** – how we let people know how they are doing in organisations. The assumption is that feedback allows people to see how they should modify their actions so that they can achieve better individual and organisational outcomes.
- **Rating and measuring performance** – the ways in which we quantify performance. Ratings and measurements are typically seen as facilitating decision-making about reward, selection and progression, as well as informing development plans and strategies.

On the following pages, we summarise the evidence in each area so that, having defined what your purpose is for performance review (or its alternatives), and before starting discussions with stakeholders about how you could alter your approach, you have practical insights on which to base your recommendations.

For each of the factors, we outline what the evidence suggest works, who it is likely to work with, and what conditions are required for it to work.



1. Goal setting

Goal setting has been established practice in organisations for a long time, and has no doubt existed in some shape or form as long as human beings have worked in any formal way. The first known empirical study of goals was by Cecil Alec Mace in 1935, and Edwin A. Locke, perhaps the most well known proponent of goal setting, began to examine the phenomena in the mid-1960s and continued researching goal setting for over thirty years.

In a recent article published in HR Magazine, Professor Rob Briner, a well-respected academic and advocate of evidence-based practice in HR, wrote defending the fundamental principles of goal setting.

“Does it work?”

You bet. Evidence for the basic principles of goal setting is strong and plentiful. Studies conducted in many settings across many occupations have replicated the basic findings. As evidence for HR practices goes it just doesn't get any better. No debunking needed. Here's a practice that actually really works.”

Of course, some goal setting won't work because it is done at the wrong time, in the wrong place, in the wrong way with the wrong people, but we can start from the basic premise that in most contexts, goal setting is useful.

The risks of goal setting

Where goals are too challenging and cannot be achieved

Many studies and writers in the field of organisational psychology challenge the idea that stretching goals are always good (e.g. Ordóñez, 2009; Karl Weick, W.E. Deming). They point to the fatiguing effects that pushing people too hard over long periods, without allowing them to reach goals with reasonable amounts of effort.

When you want people to adapt, take a broader focus and adhere to ethics

Bouskila-Yam & Kluger (2010) argue in their review of the literature that goal setting can narrow focus and actually lead to unethical behaviour. Schweitzer et al (2004) conducted a laboratory study looking at the role of goal setting in motivating unethical behaviour and demonstrated that where goals were unmet there was a particularly strong relationship, especially where individuals were just short of reaching their goals. Again this is a laboratory study so we need to be cautious about generalising to organisations but the study is interesting nonetheless.

When goals aren't set often enough

Kuvass, Buch & Dysvik (2014) demonstrated a link between people seeing their goals as unmovable and non-negotiable, and lower work performance. Interestingly this relationship is affected by how much autonomy they felt they had more generally in their job, with high levels of autonomy leading to a weaker link between the static or invariable nature of goals and job performance.

Kuvass, Buch & Dysvik suggested that this link could be managed through regular communication, adjustment and re-framing of goals, as well as allowing individuals the chance to provide an explanation as to why set goals could not be achieved due to changing circumstances.

When people don't have the levels of self-efficacy, or feelings of being in control of one's own behaviour and environment, required for sustained effort against reasonably challenging goals

We might not know enough about individuals within our organisation to make an assessment of this for ourselves, but researchers such as Albert Bandura have consistently demonstrated a link between self-efficacy and a huge variety of observable behaviours and their outcomes in the workplace. Several studies have specifically linked self-efficacy to goal attainment (e.g Zimmerman et al 1992; Shunk 1990) and this may be something we can observe happening in our own organisations. We can set people goals in similar ways; some will strive and persist, eventually achieving it while others will fall at the first hurdle, giving up easily and making excuses.

If you set challenging goals in a situation where you do not want people to take risks in their approach or exhibit unethical behaviour.

Larrick et al (2009) found that those motivated by specific challenging goals took more risks than those given a 'do your best' goal. Bear in mind, however, that the sample used in this study consisted of MBA students, raising some questions about the findings' generalisability.

The case for a strengths-based approach to goal setting

Some researchers and writers in the field argue that we can mitigate the risks of goal setting if we adopt a strengths-based approach. Whilst there is little compelling evidence of such an approach, it logically addresses the findings around when goal setting doesn't work.

A strengths-based approach involves reviewing what people are already good at and then setting goals that play to those strengths. Gallup claim that "Done correctly, a strengths-based goal setting process clarifies what the organisation means by success and whether each employee is achieving it. Employees who intentionally apply their strengths to their work increase the odds of their success. And when these employees are working at their best, they show their colleagues and managers what to expect from them."

Evidence-based warning:



Though this makes sense, no evidence is cited. Gallup are selling a strengths-based approach – are they likely to be the most neutral commentators on whether the approach works?

Summary

Goal setting can work but it's hard to do properly and the organisational context and conditions need to be right to ensure that it has a positive impact on individual performance. Goal setting is therefore not a panacea and practitioners need to remain sensitive to the context where it is being used.

2. Feedback – how much and what type?

A strong argument for the traditional performance review is that it provides a formalised opportunity for feedback. And feedback is widely accepted as a ‘good thing’ in most organisations. Using the evidence, we have taken a closer look at what type of feedback works and why the feedback given as part of performance review may not be doing what it is intended to do.

We examined the following aspects of feedback:

- Frequency and timeliness
- The nature of feedback
- How feedback is perceived
- Context

Frequency and timeliness

The ‘recency effect’ is a well-established phenomenon that describes peoples’ tendency to remember, attend to and then hypothesise, based on the information they’ve received most recently. In practice this means that, when giving feedback, managers are likely to rely on evidence and examples that are more recent, not necessarily more representative of performance. If this feedback is only given once, twice or even four times a year, it is likely to be focused on recent events rather than be a balanced view using evidence across the timeframe.

The case study looking at Adobe’s performance review identifies this and raises serious issues about the fairness of feedback given infrequently. It also notes that infrequent feedback represented a higher ‘cognitive load’ for managers, in that they simply have too much to remember if feedback is only given infrequently.

Some feel that giving feedback more regularly also mitigates the impact of giving negative feedback. The idea being that giving people feedback soon after a poor behaviour or action is observed, gives them a stronger sense of being in control of putting these behaviours and actions right. Adobe felt that negative feedback given in a timely manner was less ‘catastrophic’ for employees.

But evidence suggests that immediacy and frequency don’t work on their own. It also helps to consider the nature of that feedback and the context in which it is delivered.

The nature of feedback

Logic and professional experience tells us that feedback provokes an emotional reaction, be that positive or negative. And that emotional reaction has an impact on how the feedback is received and then acted upon. In our review, we set out to find out more about the impact of the type of feedback, how it is given, and what impact different reactions will have on the outcome of that feedback.

Negative or positive feedback?

Both anecdotal and research evidence points towards negative feedback having a detrimental effect on overall performance, as well as other measures such as attrition. But you might argue, what's the point of a performance review if we can't point out and put right behaviours that don't work for the organisation? In fact, the evidence suggests that different people react in different ways to different types of feedback, and that different ways of giving that feedback also have an effect.

Cubertson et al (2013) carried out an in-depth study with 234 staff in a university to understand the relationships between goal orientation and reactions to feedback, looking specifically at the impact of individual goal orientations on that relationship.

Overall their findings are complex and suggest that one size does not fit all when it comes to feedback. Some people prefer negative feedback (if it comes with developmental advice and information) while others will have a strong negative reaction to negative feedback, which appears to reduce their orientation towards certain types of goals.

Another interesting study looked at the impact of how negative feedback is given. Asmuß (2008) found that if negative feedback is phrased as 'socially unproblematic' then it is discussed briefly and directly with few problems. Where the opposite is true (i.e. the issue is presented as socially unacceptable), then the negativity is focused upon and discussions and ruminations tend to be lengthier.

Self-appraisal

In our review of the literature, we identified several studies that demonstrated that feedback that includes an element of self-appraisal tends to have a more positive impact on performance and perception.

Asmuß (2008) found that negative feedback generated by both employer and employee tended to result in 'healthier' discussions and action-oriented discussions. This is a finding replicated in other studies. For instance Ramaswamy & Thanga (2014) conducted a study with 115 employees in Saudi Arabia and found that self-appraisal resulted in people perceiving higher accuracy, fairness and improved understanding of the demands and expectations placed on them. However, based on the sample included in this study, practitioners in the UK may want to consider how generalisable the results are.

Evidence-based warning:



Can we generalise these findings given where they were conducted?

How feedback is perceived

If the nature of feedback is important, then how it's perceived also plays a critical role. Interpretations of the purpose and intention behind feedback will lead to different responses and therefore different behavioural outcomes.

Research evidence (e.g. Kuvaas, Buch & Dysvik, in press) suggests that perceived 'constructiveness' of supervisor performance and work performance have a positive relationship, but only when feedback is also perceived to be immediate and frequent.

Constructiveness in this study meant 'strengths-based feedback' – i.e. feedback that focused on the skills and behaviours that contributed towards good performance, task-related rather than of a personal nature, and that it was perceived as acceptable and accurate.

David (2013) also supported this idea with her study of Year 1 and Year 2 clinical nurses working in a hospital. She suggested that the extent to which a manager treats (or is perceived as treating) an employee in a dignified, constructive and non-harsh manner is the most important factor influencing how receptive he or she will be to feedback. And these perceptions of how well they had been treated were linked to more positive emotions, which in turn linked to higher performance ratings in the following year. This study is limited in its utility however, as it is based on correlational relationships – it doesn't demonstrate causality.

Context

Again this might seem like common sense, but there is evidence to suggest that feedback works best when the organisation provides a context in which employees can develop skills, as well as an atmosphere that signals that it is possible for them to do so, and that individuals have the capability for improvement and development (Maurer, 2002).

**Evidence-based
warning:**



Correlation doesn't mean
causality.

Summary

So when it comes to feedback, what can we conclude?

- Managers may be subject to the 'recency effect' and over rely on recent evidence, rather than a more representative sample of performance behaviour over a period.
- Immediacy and frequency are two factors which can improve the quality and accuracy of feedback, but we also need to consider the 'tone' of the feedback – whether it is largely positive or negative and how constructive it appears to the recipient.
- This all points to the need for managers to stay on top of performance conversations, using an appropriate style and representative evidence, not just what is available from recent memory.
- The organisational context in which the feedback is provided is also an important consideration. Contexts that provide opportunities for employees to develop, and support the belief that improvement is possible, contribute to the overall positive impact of feedback.

3. Rating performance

Many studies have highlighted perhaps the biggest problem with performance review – that generally, people are really bad at accurately and fairly assessing the performance of others. Further, employees are far more concerned by their rating, than they are in attending to feedback, especially when related to pay.

Given that the central function of performance review is usually to rate performance in a way that allows comparison and leads to action (e.g. pay, promotion, exit) this should be a very good reason for trying something different. But before advocating dispensing with assessment, measurement and rating, we had a more detailed look at the evidence to see what is really going on when people rate others.

The rating scale

The use of ratings in performance appraisals is deeply rooted in the history of personnel psychology (Landy & Farr, 1980 in Brutus 2010). The approach an organisation uses varies according to what they are using the information for, and range from 5-point scales to stack ranking, where the workforce is segmented into three categories, with the lowest 10% being at risk of dismissal.

Stack ranking (pioneered by Jack Welch of General Electric) was, by 2012, used by 60% of Fortune 500 firms, but is being increasingly abandoned by organisations (e.g. Microsoft) wishing to project a more progressive image. Stack ranking is increasingly seen as the product of an age where employers were viewed strictly as 'workers' and where performance could be easily measured in terms of outputs.

In studying the effect of stack ranking at Microsoft, Deloitte found that the ranking system resulted in "capricious rankings, power struggles among managers, and unhealthy competition among colleagues."

Research has demonstrated that there are pros and cons for different types of rating scales, depending upon their intended use. A particularly extensive study by Weijters et al (2010) generated a somewhat complicated decision framework for selecting a rating scale for a wide variety of purposes. Their work suggests that to overcome biases and to ensure appropriate response distributions, only a 5 or 7-point 'fully labelled' scale will be sufficient in most cases.

It would appear that a 3-point scale does not sufficiently differentiate, and even though a 5-point scale tempts people into central tendency (rating too many people as a neutral '3'), the distribution of responses tends to reflect expected distributions of performance more accurately.

Some organisations, where a more focused performance measurement is required, favour an even numbered scale to resist this, but this can lead to a winners and losers perception of performance rating and a general tendency for people to rate over-favourably.

And of course, scales only work if they are properly defined and labelled, they measure observable behaviours and outputs, and that everyone understands them.

Bias

A good deal of research has demonstrated that unconscious, and sometimes conscious, biases play a part in how people rate others.

Scullen et al (2000) attempted to quantify the extent to which these biases influence performance ratings in a large sample using a multi-rater generic feedback instrument. They found that a greater proportion of the variance in ratings was down to the biases of the rater than the performance of the person. Other studies have found that rater liking of the individual has a stronger impact on the rating of performance than actual performance (Sutton et al, 2013). In other words, it's not what you do that really matters, but whether you are liked.

Remember that correlation doesn't mean causation. We could interpret these findings either as a sign that we should dispense with ratings completely or that perhaps there is something else going on here that needs investigating. Perhaps the rating scale itself is insufficiently detailed or poorly designed, leading to inaccuracy (see above), or perhaps the raters just need to be better trained (see below).

Training and definition

One of the themes we explored in our review of the literature was whether training for raters and more structured approaches to defining the criteria for ratings, made an impact on the accuracy and fairness of ratings. The answer seems to be yes, not least because people report that they do not feel comfortable with approaches they don't understand.

In the early 1980s, there was a shift from training raters to avoid biases such as 'halo', leniency and other common rating errors to more proactive rater accuracy approaches. This included the development of new rater training programmes that aim to increase rating accuracy, and included proposed Frame Of Reference (FOR) training, Behavioural Observation Training (BOT), and training familiarising raters with Behaviourally Anchored Rating Scales (BARS).

Frame of Reference (FOR) Training: Training raters in the metric and criteria used in a rating scale. The aim is to mitigate idiosyncrasies caused by raters whose ideas about what is important when rating performance may differ from the organisation's standards.

Behavioural Observation Training (BOT): Training focused on how raters observe behaviour rather than how they evaluate it. The training aims to improve how people observe, recall and process their observations and often centre around the ORCE model (Observe, Record, Classify, Evaluate) which is used in much assessment centre training.

Behaviourally Anchored Rating Scales (BARS): Training raters in how to use the BARS in performance appraisal. BARS are scales where each point is a numerical rating that is described in terms of examples of likely observed behaviours.

Two Canadian studies explored this in more detail. Noonan & Sulsky (2001) conducted a study that found that managers in the Canadian Air Force made more accurate ratings after Frame of Reference training but Behavioural Observation training had little effect.

Catano et al. (2007) focused on the utility of training against a Behaviourally Anchored Rating Scale in the Royal Canadian Mounted Police (RCMP). They concluded that basing the competency profile within an appraisal system on competency-related behaviours can increase both the reliability and the validity of the inferences drawn from those profiles.

We can also borrow evidence from the world of assessment centres.

The British Psychological Society has recently reviewed its standards in this area, and the evidence they collected led to a recommendation that all assessors need training in the criteria used and in the ORCE approach (Observe, Record, Classify and Evaluate). This recommendation, and the evidence that supports it, may lead us to conclude that it is sensible to ensure that training on any rating scale is needed and that we need to work out the most efficient and cost effective way to do this, rather than not do it.

Evidence-based warning:



Can we generalise these findings given where they were conducted?

Summary

When it comes to ratings, it is difficult to draw concrete conclusions. The evidence suggests that organisations need to be exceptionally cautious about the approach they use.

- The process of rating others seems to be problematic, because people find it difficult to rate others accurately and because people can react badly to anything other than a positive rating.
- Training can help people rate more accurately, but the modest gains in rating accuracy may call this investment into question.
- Your rating scale needs to be carefully chosen to suit your purpose. If you aren't clear about why you are rating, then you may be choosing the wrong method. For performance review processes in the main, a 5-point labelled scale looks like the approach most likely to generate more accurate and reliable ratings.

Returning to the systems perspective

Our examination of the literature on this topic points to the need to view the performance review in a holistic way. That is, there are important factors at play at a number of levels:

- **Intra-personal level:** Factors such as rater and ratee personality and ability, preferences for feedback type and experience of feedback to date.
- **Inter-personal level:** Relevant factors include the relationship between managers and their direct reports, the role of bias and expectations in the performance review process.
- **Organisational level:** It's key to take account of organisational culture, whether the environment supports growth and development and whether the organisation wants to use the review for development, ranking of employee performance or simply to allow managers and team members to 'keep in touch'.

As such, there's no 'one size fits all' advice when it comes to performance reviews. However, that's not a negative or disappointing message. What it means is that organisations should look to their own environment and make informed decisions about how they want to run these reviews and rate themselves in terms of factors explored above. Looking to other organisations may not be the best approach – their context, culture and aims could differ significantly from your own.

The other key point to bear in mind when taking a holistic approach is the fact that your organisation is a complex system. Making adjustments to something as central as the performance review could have unintended negative consequences elsewhere in the system.

By removing your performance review, you may unintentionally impact employee motivation and productivity, and damage your employer brand to prospective applicants. Changes to such a central process within the organisational system shouldn't be undertaken from a single perspective.

The performance review can impact employee performance, manager-subordinate relationships and turnover intentions – positively or negatively.

Some key questions to consider

Before making changes to something so important, we recommend you take the time to reflect and discuss these key questions:

- How would you rate the success of your performance review process? Is this based on data and evidence of your own personal experience?
- Are you clear on the goals of your performance reviews? Are you confident that these goals are clearly understood by all stakeholders?
- Have you got the right sort of work and organisational culture to support good quality goal setting?
- Is the work straightforward and outcome focused, or is it complex or intellectual with unclear outcomes?
- Does the nature of your work lend itself to fixed annual or bi-annual goals or should you consider a more flexible approach, perhaps setting short-term, negotiated and flexible goals?
- What are managers' and staff's attitudes towards your current approach to performance review? Do you have a communications channel that allows them to give you feedback on this topic?
- How many hours annually are spent on performance reviews? Given that investment, are you sure that you are getting any sort of return?
- Have you provided training on your performance appraisal system, particularly in relation to goal setting and rating?
- Do your employees have access to resources (e.g. training and wider development activities) that allow them to improve their performance?
- Do your managers provide constructive feedback that is also immediate and frequent?
- Have you considered how you can evaluate the effectiveness of your performance review process?



How can we help?

Understanding where you are now

If, as an organisation, you're not sure what value your existing performance review process is adding, we can help you evaluate the process and identify its strengths and areas for improvement.

Improving your current process

If you'd like to improve the quality of your performance review, we can design and run appropriate training for your managers.

Designing and evaluating a new performance review process

If you're considering abandoning the performance review completely, we can evaluate your plans with you, and help you design and implement an alternative process which meets the needs and culture of your organisation.

Get in touch

To discuss how we can help with your performance review process,

call us on

020 7947 4273

or write to

info@futureworkcentre.com

References

- Asmuß, B. (2008). Performance Appraisal Interviews, Preference Organization in Assessment Sequences. *Journal of Business Communication*, 45(4), 408-429.
- Asplund, J. & Blacksmith, N. (March 2012). Strengths-Based Goal Setting. Retrieved February 15, 2016, from <http://www.gallup.com/businessjournal/152981/strengths-based-goal-setting.aspx>
- Bouskila-Yam & Kluger (2010) Strength-based Performance Appraisal and Goal Setting, *Human Resource Management Review* 21, 137-147.
- Briner, R. (January 2016). What's the evidence for...Goal Setting. Retrieved February 15, 2016 from <http://www.hrmagazine.co.uk/article-details/whats-the-evidence-for-goal-setting>
- British Psychological Society & the Division of Occupational Psychology (2015) The Design and Delivery of Assessment Centres. Retrieved February 15, 2016 from https://www.bps.org.uk/system/files/user-files/Division%20of%20Occupational%20Psychology/public/inf234_assess_centres_final.pdf
- Brutus, S. (2010). Words Versus Numbers: A Theoretical Exploration of Giving and Receiving Narrative Comments in Performance Appraisal. *Human Resource Management Review*, 20(2), 144-157.
- Buckingham, M., & Goodhall, A. (2015, April 1). Reinventing Performance Management. Retrieved January 12, 2016, from <https://hbr.org/2015/04/reinventing-performance-management>
- Catano, V. M., Darr, W., & Campbell, C. A. (2007). Performance appraisal of behavior-based competencies: A reliable and valid procedure. *Personnel Psychology*, 60, 201-230.
- Culbertson, S. S., Henning, J. B., & Payne, S. C. (2013). Performance Appraisal Satisfaction. *Journal of Personnel Psychology*, 12(4), 189-195.
- David, E. M. (2013). Examining the Role of Narrative Performance Appraisal Comments on Performance. *Human Performance*, 26(5), 430-450.
- Gahan, P. (2015) Don't throw out the baby with the bathwater on performance reviews. Retrieved February 15, 2016 from <http://www.smh.com.au/business/comment-and-analysis/dont-throw-out-the-baby-with-the-bathwater-on-performance-reviews-20150730-giniod.html>
- Garr, S. & Liakopoulos, A. (2014), Deloitte Global Human Capital Trends 2014: Engaging the 21st-century workforce. Retrieved February 15, 2016 from <http://www2.deloitte.com/global/en/pages/human-capital/articles/performance-mgmt-is-broken.html>

- Hinds, R., Sutton, R., & Rao, H. (2014). Adobe: Building Momentum by Abandoning Annual Performance Reviews for "Check-ins". Stanford Graduate Review of Business. CASE: HR-38. Retrieved February 15, 2016 from <https://www.gsb.stanford.edu/faculty-research/case-studies/adobe-building-momentum-abandoning-annual-performance-reviews-check>
- Kuvaas, B., Buch, R., Dysvik, A. (2014). Performance Management: Perceiving Goals as Invariable and Implications for Perceived Job Autonomy and Work Performance. *Human Resource Management*, 51(1) doi: 10.1002/hrm.21680
- Kuvaas, B. Buch, R. & Dysvik, A. (in press). Constructive Supervisor Feedback is Not Sufficient: Immediacy and Frequency is Essential. *Human Resource Management*.
- Larrick, R. P., Heath, C., & Wu, G. (2009). Goal-Induced Risk Taking in Negotiation and Decision Making. *Social Cognition*, 27(3), 342–364.
- Maley, J. (2013), "Hybrid purposes of performance appraisal in a crisis", *Journal of Management Development*, 32(10), 1093 – 1112.
- Maurer T.J., Mitchell, D.R.D., Barbeite, F.G. (2002), Predictors of Attitudes towards a 360-degree feedback system and involvement in post-feedback management development activity. *Journal of Organizational Psychology*, 75, 87-107.
- Noonan, L. E., & Sulsky, L. M. (2001). Impact of Frame-of-Reference and Behavioral Observation Training on Alternative Training Effectiveness Criteria in a Canadian Military Sample. *Human Performance*, 14(1), 3–26.
- Ordóñez, L. D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009). Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting. *Academy of Management Perspectives*, 23(1), 6–16.
- Ramaswamy & Thanga (2014) Self Performance Appraisals and its Effectiveness in Performance Management; an Empirical Study in Saudi Arabia, *An International Journal of Management*, 5.
- Rock, D., Davis, J., & Jones, B. (August 2014). Kill Your Performance Ratings. Retrieved January 11, 2016, from <http://www.strategy-business.com/article/00275?gko=c442b>
- Scullen, S., Mount, M., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*. 85(6), 956-970.
- Sutton, a W., Baldwin, S. P., Wood, L., & Hoffman, B. J. (2013). A Meta-Analysis of the Relationship Between Rater Liking and Performance Ratings. *Human Performance*, 26(5), 409–429.

About the Future Work Centre

Our mission is to make work better for everyone, now and in the future. We want to examine what evidence there is for the advice that people and organisations are offered about work. And we want to know whether that advice works. Put simply, we want to find out what works, in what way, and for whom.

To do this we:

- **Educate** organisations, the public and early career psychologists about the science of occupational psychology and its application in the workplace.
- Use innovative psychological **insight** to shine a light on how work affects people, organisations and society – and what does and doesn't make it better.
- Deliver high quality, independent occupational psychology **services** to organisations, helping them make better decisions and investments.

Find out more about the Future Work Centre at
www.futureworkcentre.com

 020 7947 4273

 info@futureworkcentre.com

 @FW_Centre

