Performance appraisal and true performance

Mark Cook

The author

Mark Cook is a Chartered Occupational Psychologist with the Personnel Selection Research Group at University College, Swansea, UK.

Abstract

Argues that the conventional validation paradigm, which uses subjective performance or appraisal ratings as criteria, may be of doubtful validity. Discusses research into performance appraisal which documents four sets of problems which may reduce the usefulness of performance ratings as criteria. These problems include biases, politicking, impression management and undeserved reputation. Describes the inaccuracies to which these problems give rise and concludes that instead of selecting the right people for management, selection methods validated against appraisal will simply perpetuate an unsatisfactory status quo.

Performance appraisal and performance-related pay are increasingly popular in the UK at present, and both seem to be viewed with great favour by the present Government. In most sectors appraisal is introduced first at management level, and then "cascades" down to less-exalted levels of staff. Before this happens, it might be timely to review research on performance appraisal which documents the many problems of performance ratings (PRs), and casts serious doubt on their validity as measures of true work performance.

Biases

PRs suffer from many biases, meaning factors that "should not" influence raters but do, or factors that influence ratings in ways about which raters are unaware.

Age, ethnicity, and gender

The pooled results of 40 separate researches[1] find a low *negative* correlation between age and PRs, whereas age and objective indices of performance correlate *positively*. This implies that PRs show substantial bias against older persons.

Similar large-scale analysis[2] finds a small but consistent "own race" bias in PRs; whites favour whites, blacks favour blacks, etc.

Gender also distorts PRs. Sometimes PRs show simple gender bias[3], so that women with masculine characteristics are rated more promotable. Sometimes the bias is subtler, involving gender stereotype of the occupation, so that raters who do not see management as a traditionally female occupation tend to give women managers lower PRs[4].

Physical appearance

Other biases in PRs are subtler. "Common sense" argues that physical attractiveness is not likely to influence PRs, because "beauty is in the eye of the beholder" and anyway should not influence behaviour in the workplace. In fact there is a very reliable broad consensus about who is and who is not good-looking, and strong evidence that appearance does affect PRs. Two separate researches[5,6] find a strong link between attractiveness – rated on a five-point scale – and later salary level. Both find that each extra scale point of rated attractiveness is "worth" $2,000-2,600 more in salary. This implies that the salary difference between the most and least attractive could be as great as $13,000. The effect is far stronger

for males, in "male" jobs, and in older age groups. PRs are strongly biased by appearance, and the bias is not confined, as "common sense" might argue, to young female ratees. In fact it is probably strongest in traditional management circles.

Attitudes and values

Nursing managers' PRs are shaped more by the manager's belief system than by the subordinate's actual work performance[7]. The research uses a "laddering" technique to explore nursing managers' concepts of effectiveness. This means continuing to ask managers the question "why is that important" until one eventually elicits answers like "self-fulfilment" or "pride in being a good manager", which the researchers argue represent the managers' fundamental values. The managers *said* that staff work behaviour shaped their PRs, but "policy capturing" analysis showed that the managers' own values were the best predictor of their PRs. Subordinates who helped managers to feel proud of being a good manager, or helped to make them feel fulfilled, received better PRs.

In-group/out-group

One school of thought argues that most work groups can be divided into an in-group, whose members enjoy the supervisor's confidence and concern and get assigned the more challenging tasks, and an out-group, whose members are treated like "hired hands" and are assigned the mundane tasks. Research on bank staff[8] shows that in-group members get better PRs but do not perform any better on objective performance indices. This implies that the in-group members achieve their position not by better work, but by some other path.

Personal like/dislike

"In-depth" interviews with managers who regularly make PRs reveal that three-quarters freely admit they allow liking to inflate PRs and dislike to lower PRs[9]. The same analysis also reports that 83 per cent of managers say being in a good or bad mood shifts the PRs they make – probably downwards in most cases, given that 73 per cent say they hate having to make PRs.

PRs can be analysed to detect age, race or gender bias (and usually are, to avoid equal opportunities problems). PR data cannot be analysed to detect bias created by appearance,

like/dislike, similarity of outlook, or in-group/out-group effects, because this information is not routinely recorded, and often it is not readily accessible.

This means that many sources of bias render performance appraisal ratings less accurate, but in ways which neither employer nor employee can easily identify. In technical terms, the performance ratings are *contaminated*, and the contamination is *invisible*.

This in turn has further implications for research on personnel selection. Between 60 per cent and 80 per cent of researches on selection use supervisor rating as their criterion of "true performance". Bias in ratings, from the sources described above, means that rating may often be a fairly poor criterion of true performance, which in turn implies that much research may actually underestimate the true validity of selection methods.

Politicking

Organizations usually assume that managers *want* to make accurate PRs, and are prevented only by unconscious biases which can be cured by sufficient training or the right rating format. The in-depth interview research flatly contradicts this and concludes that "political" considerations, or "private agendas", nearly always influence PRs. The managers list many reasons for giving *inflated* PRs: to maximize merit increases, to protect persons whose performance is suffering because of personal problems, to avoid "washing dirty laundry in public", to avoid creating a badly written record, to avoid confrontation, and to rid themselves of people by promoting "up and out".

The interviews also uncover managers' reasons for making deliberately *low* PRs: to shock someone back onto a higher "performance track", to teach a rebellious subordinate who is in charge, to send someone a message that they should consider leaving the organization, and to build a well-documented record of poor performance to speed up termination.

Gross leniency in appraisal rating is very apparent, but subtler "politicking" may not be. Like biases based on appearance or attitude, subtler politicking contaminates the appraisal rating, and makes it less accurate, both as a measure of performance, and as a criterion in research of selection methods.

Impression management

Many organizations set great store by relatively trivial details, that impress senior management but have little to do with effective work performance: the right mannerisms, the right clothes, or the right buzzwords. Padfield's history[10] of battleships describes how promotion in the Royal Navy of the 1890s went to officers with the most highly polished ships, which caused a few to behave as if they had forgotten what battleships were for; they avoided gunnery practice in case the powder smoke spoiled their paintwork. An extreme case of this trend may be termed the World War I mentality. Organizations occasionally exist in which subordinates gain credit for pushing ahead with management plans that are absurdly wrong, in pursuit of aims which are completely pointless, stifling criticism either of purpose or of method with cries of "commitment" and "loyalty".

Ingratiation

English has a rich vocabulary to describe workplace ingratiators – including words listed in dictionaries as "not in polite use" – which implies that the behaviour is widely recognized, but not widely popular. One research[11] identified three underlying types of ingratiating behaviour, or "upwards influence styles":

(1) *Job-focused ingratiation*: claiming credit for things you have done, claiming credit for things you have not done, claiming credit for what the group has done, arriving at work early to look good, working late to look good.

(2) *Supervisor-focused ingratiation*: taking an interest in the supervisor's private life, praising the supervisor, doing favours for the supervisor, volunteering to help the supervisor, complimenting the supervisor on his/her appearance and dress, agreeing with the supervisor's ideas.

(3) *Self-focused ingratiation*: presenting self to the supervisor as a polite and friendly person, working hard when results will be seen by the supervisor, letting the supervisor know that you are trying to do a good job.

Research suggests however that ingratiation does not always succeed in obtaining good PRs. Unsubtle ingratiation may sometimes be too blatant to be credible, or palatable.

Ingratiation and other impression management techniques also contaminate appraisal ratings, and make them less accurate reflectors of true worth to the organization. Besides undermining performance appraisal, and selection research, this tends to be bad for morale, when staff see persons whose true performance is poor, but who are good at ingratiating themselves, get merit awards, or promotion, or other marks of favour.

Undeserved reputation

A good reputation can be earned by good work, but many features of large organizations make it possible to earn one in other ways.

It's who you know, not what you know

A widely voiced observation, which implies that the employee's time may often be as usefully spent creating a network of allies and contacts as in doing any actual work. "Team building" or being "a team player"' are phrases often used in job advertisements and personnel specifications. Co-operating with colleagues is self-evidently important in most work – but could "team-player" sometimes also refer to people who are very good at making themselves liked or good at appearing to be indispensable, but who are not necessarily very effective performers? A closely knit "team" is also more likely to share, and stick closely to, a *social reality* (see "Who sets the standards", below).

The non-working day

Research on US army infantry[12] showed that the average soldier spends 25 minutes of a typical "training day" receiving instruction and 57 minutes practising the task; the rest of the day is swallowed up by parades, being "en route", polishing boots, hanging about, etc. Similar research[13] shows that a substantial part of the manager's day is spent doing things that cannot be unambiguously linked to the accomplishment of specific tasks.

Polishing the image

Some individuals consciously set about building themselves a reputation, by conventional "public relations" techniques of controlling information about themselves and their doings.

Cover your back

In cautious organizations a good reputation is built largely by *not* doing things: not making controversial decisions, not attracting complaints, not getting bad publicity.

Attributability

Complex organizations and/or long time scales mean that it is often difficult to assign true responsibility for outcomes. When it is difficult to know who is really responsible for successes and failures, it is correspondingly easier for devious individuals to steal the credit for successes, avoid the blame for failures, and build an undeserved reputation.

"Kicking over beehives"

In some sectors people move jobs every year or so, so that they can make many changes but escape before any consequent problems emerge. Managers call this "kicking over beehives". Rapid turnover also makes it difficult to know who is responsible for what.

Who sets the standards

A company that manufactures gearboxes has its success defined externally, and unambiguously, by its sales figures. A university by contrast constructs its own *social reality* about success. The academic staff decide what issues are worth researching, and what subjects should be taught and by implication whose work has merit. Where success is defined by the organization and its staff, greater scope exists for creating undeserved reputations.

Empire-building

In many organizations, success consists of increasing the size of one's department or budget. Services are provided for the sake of justifying the organization's expansion. Cyril Northcote Parkinson[14] describes how staff levels in the British Colonial Office rose steadily throughout the twentieth century, while the number of British colonies fell equally steadily.

Reorganizations

Reorganizations, large or small, create a perfect form of pseudo-work, divorced from external standards. Everyone's efforts are centred for months on something that has no end product, and frequently serves no useful purpose, but is an ideal environment for the person who seeks to build a reputation.

Reorganizations also make it easy to blur responsibility for successes and failures.

Making your mark

It has been said of Parliament that no Minister has really "arrived" until he/she has passed an Act. Critics argue that this explains the enactment of numerous badly drafted and unenforceable laws. Similar "mark making" mechanisms include opening a new department or new building (or often, these days, *closing* one), and reorganizations (see preceding paragraph). Often the more outrageous the mark making, the better. Even now, practically everyone remembers the name of Dr Richard Beeching, if not as the author of *Reshaping the Railways*, then as a code name for someone who was brought in to close things down.

Some job advertisements appear to be groping towards aspects of this "mark making" mechanism with references to "stature" (not referring to physical build) and "credibility" (not meaning that the person has not been caught telling lies recently).

Implications for PRs

Someone whose true performance is poor, but who is a good self-publicist, or good at claiming responsibility for others' work, or successful at defining his/her achievements as useful and important, will achieve better PRs. Conversely someone whose true performance is good, but who is not a good self-publicist, or does not claim responsibility for others' successes, or who fails to get his/her work defined as important, may get poorer PRs. The PR has been contaminated, and less accurately reflects true performance.

In blue-collar jobs there is often a simple objective criterion of performance, of the units-produced-or-processed type. When these objective measures of work performance are compared with subjective appraisals for the same employee, a very low correlation results ($p = 0.27$)[15], which means that objective measures and the subjective appraisals are not measuring the same thing. Perhaps the same is true of management, which implies that those who get the good ratings, and the promotions, and the performance-related pay, may not always, or even often, be the ones who are "really" doing a good job.

This has a further implication. Most selection research uses performance rating as its criterion of "true performance". If performance appraisal is *not* a good measure of true performance, then selection methods validated against appraisal will not be selecting the "right" people for management, but will simply be perpetuating an unsatisfactory status quo.

References

1 Waldman, D.A. and Avolio, B.J., "A meta-analysis of age differences in job performance", *Journal of Applied Psychology*, Vol. 71, 1986, pp. 333-8.

2 Kraiger, K. and Ford, J.K., "A meta-analysis of ratee race effects in performance ratings", *Journal of Applied Psychology*, Vol. 70, 1985, pp. 56-65.

3 Hartman, S.J., Griffieth, R.W., Crino, M.D. and Harris, O.J., "Gender-based influences: the promotion recommendation", *Sex Roles*, Vol. 25, 1991, pp. 285-300.

4 Landy, F.J. and Farr, J.L., "Performance rating", *Psychological Bulletin*, Vol. 87, 1980, pp. 72-107.

5 Frieze, I.H., Olson, J.E. and Russell, J., "Attractiveness and income for men and women in management", *Journal of Applied Social Psychology*, Vol. 21, 1991, pp. 1039-57.

6 Roszell, P., Kennedy, D. and Grabb, E., "Physical attractiveness and income attainment among Canadians", *Journal of Psychology*, Vol. 123, 1989, pp. 547-59.

7 Jolly, J.P., Reynolds, T.J. and Slocum, J.W., "Application of the means-end theoretic for understanding the cognitive bases of performance appraisal", *Organizational Behavior and Human Decision Processes*, Vol. 41, 1988, pp. 153-79.

8 Vecchio, R.P. and Gobdel, B.C., "The vertical dyadic linkage mode of leadership: problems and prospects", *Organizational Behavior and Human Performance*, Vol. 34, 1984, pp. 5-20.

9 Longenecker, C.O., Sims, H.P. and Gioia, D.A., "Behind the mask: the politics of employee appraisal", *Academy of Management Executive*, Vol. 1, 1987, pp. 183-93.

10 Padfield, P., *The Battleship Era*, Hart-Davis, London, 1972.

11 Wayne, S.J. and Ferris, G.R., "Influence tactic affect, and exchange quality in supervisor-subordinate interactions: a laboratory and field study", *Journal of Applied Psychology*, Vol. 75, 1990, pp. 487-99.

12 Bialek, H., Zapf, D. and McGuire, W., *Personnel Turbulence and Time Utililization in an Infantry Division*, HumRRO FR-WD-CA.77-I I, Human Relations Research Organization, Alexandria, VA, 1977.

13 Campbell, J.P., Dunnette, M.D., Lawler E.E. and Weick K.E., *Managerial Behavior, Performance, and Effectiveness*, McGraw-Hill, New York, NY, 1970.

14 Parkinson, C.N., *Parkinson's Law*, John Murray, London, 1958.

15 Heneman, R.L., "The relationship between supervisory ratings and results-oriented measures of performance: a meta-analysis", *Personnel Psychology*, Vol. 39, 1986, pp. 811-26.