

UPWARD APPRAISAL: PARTICIPATIVE MANAGEMENT OR WINDOW DRESSING?

Upward appraisal has for some time now been regarded as a strategy for creating the involvement of front-line employees within the workplace, and it is estimated 10%-15% of today's organisations use either upward appraisal or 360 degree feedback for developmental learning and performance evaluation, (Yammarino & Atwater 1997). The claims are that the process creates a perceptual convergence and a clarity of communication concerning important organisational values and behaviours across hierarchies. This leads to slicker team working and enhanced managerial effectiveness. However psychometric properties for upward appraisals have consistently returned one factor, halo type solutions and poor interater reliability: 0.215 (Mount, K. 1984), 0.18 (London & Wohlers, 1991). One possible reason for these poor properties could be that often upward appraisal questionnaires are prepared by senior/middle managers whose perceptual and cognitive categories concerning organizational behaviour and effectiveness are different from those of the front-line. This means the items are to a great extent unintelligible to the raters. The purpose of this dissertation was to see if by creating a behaviourally anchored rating based upon critical incidents gleaned from the shop floor of 3 depots of a large international company, and by using the shop floor vernacular to express the items, the psychometric properties and the utility of such an instrument would improve.

QUESTIONNAIRE DESIGN.

Over a period of one 1 month front-line employees were asked to furnish the researcher with a **specific** example of when their manager had acted efficiently, and another of when their manager had acted in an inefficient manner. The questioning was rotated at each interview to prevent a bias towards either efficient / inefficient examples. Each interview lasted about 10 minutes and 225 incidents were collected. 182 (80.9%) were of inefficient examples, and 43 (19.1%) were representative of efficient examples. As many of the critical incidents were retranslated into questionnaire items as possible. This was to ensure the resultant questionnaire was representative of the frontline mindset rather than have a situation where many of the critical incidents were dropped because they did not fit into an already existing competency framework. A table was created where each of the critical incidents were summarised and the designated question designed to measure such behaviour was put alongside. So for example one critical incident went "I injured my Achilles tendon at work. I went to my Line manager to ask him if I could put it in the accident book, and he told me he did not have the time. I did not think it was good management because apart from Health & Safety it gives the impression he does not give a toss about my welfare". The question numbers 7, and 8, were put along side this incident. This was so interested parties could trace the logic and the relevance for themselves concerning the particular items.

4 broad dimensions seemed to arise; interpersonal skills, job knowledge skills, man management skills, and a general systems dimension. The systems dimension consisted of examples that related to how the organisation was structured. There were many more incidents that related to interpersonal skills, and man management skills, than job

knowledge skills, and systems incidents were by far the least frequent. This weighting was represented in the final questionnaire.

The pilot questionnaire had 30 items. This was distributed to 40 employees, with 37 returned. On a seven point scale 47.9% of the responses fell into the typical response sets of 1, 4, or 7, meaning over 50% of responses were not typical response sets. Split half reliability was 0.94 (Pearsons).

When constructing the final 36 item questionnaire comments from the frontline were taken into account. Most issues were around clarity. Bold type was put in to each questionnaire item to highlight the main thrust of the question. Certain questions were rephrased, and finally it was noticed that some critical incidents were not sufficiently represented in the questionnaire, and the extra 6 items redressed that fact.

To give a tone of the questionnaire these are examples of one question from each dimension, drawn from critical incidents. The man management question is question 7 drawn up from incidents like the one mentioned above.

Man management:

When you bring **Health & Safety** issues to the attention of your manager does he/she take **immediate** action?

Interpersonal:

Does your manager have the ability to **assert** him/her self in a professional way when needed?

Job knowledge:

Do you find your colleagues **know more about the job** than your manager?

System:

Are the communications you get from your manager consistent?

IMPLEMENTATION.

The questionnaires were printed up at the company's divisional headquarters. The questionnaires were then administered by 20 Line managers to their teams of approximately 10 staff. This process took place in work time, and 20 minutes was allowed. The questionnaires were then collected by a frontline employee, with the Line manager's self rating being tagged for identification purposes. These were then sent in a sealed envelope for analysis, and eventual feedback during team briefs at a later date by myself and the Line manager.

RESULTS.

Even though some teams returned quite respectable coefficients for inter-rater reliability; .49 and .34 (spearman) for 2 teams, very low coefficients for other teams brought the mean down to .218 (spearman). These are similar to other findings which has looked at agreement within a rater source. (Mount, K 1984; .215 and London&Wohlers 1991 .18). Convergent validity between front-line and manager was slightly greater, with a mean of .296 (range .07-.6) Spearman. These again are similar to those reported by Mount '84 and London&Wohlers '91 .23 and .24 respectively.

Exploratory factor analysis, (principle components, with varimax rotation) returned 6 factors explaining 75.7% of the variance for the front-line employees, and 8 factors explaining 97.4% of the variance for the line managers. Examination of the scree plots however demonstrates one factor solutions for both sets of ratings, with slightly more discrimination in the ratings of the managers.

The items which loaded into these factors were totally different for the two cohorts of respondents, which is in line with the finding of Foti (1990) that the mindset of employees from differing levels from within the same organisation will be different. Interpersonal skills were often represented in the managerial factors, whereas job knowledge items loaded into the factors accounting for most variance for the frontline.

DISCUSSION.

It would appear that despite the tailoring of a questionnaire to meet the requirements of the frontline psychometric properties are still relatively poor. However what became clear in the process of this research was that there are many reasons why a good instrument would have poor psychometric properties, and listening to the feed back from the participants provided some of these reasons:

1. Many frontline felt their manager behaved differently to different members of the team, and this would account for a lack of factor structure in the final analysis.
2. No training was given to the frontline on how to effectively rate managers. Given that only 36% of variance is shared when supervisors rate, (Cook 1993), it is not surprising frontline who have no training do not rate as accurately.
3. Many frontline felt that to rate their manager honestly would incur the manager's wrath upon them. Anonymity they felt was no protection. They felt projection was at work, and the manager would "know" anyway who rated them down, and react against that employee. As a result ratings were very similar to the official upward appraisal (77.4% of the maximum possible rating compared with the official 71.33%.) However the official employee opinion survey, the largest in the UK, returned only 51% of the maximum possible rating for the similar managerial dimensions. It is feasible to suggest the inflation of the upward appraisal ratings is a function of perceived retribution by a local management who is perceived as defensive and vindictive.

4. Many employees took the “what’s in it for me?” Line. They felt they were being asked to provide information for no reward. This was reflected by the fact that only 39.9% of the questionnaires were returned. Of the 220 that were returned it was quite clear there was little involvement as many demonstrated no variance across the 36 items.
5. It was felt by many participants the whole process was flawed as sufficient time and energy was not given to ensuring that managers would produce time bound and measurable action plans based upon any feedback given by the frontline.
6. Many frontline did not want to rate accurately as they felt they obtained much more benefit by working under an inefficient manager. They could in these circumstances create overtime, enjoy greater autonomy, prolong breaks, create extra breaks, and generally benefit from a more relaxed working environment.

CONCLUSION.

In order to rate accurately one must observe the behaviour of the ratee, form a cognitive representation of that behaviour, store the representation in memory, retrieve the information needed for evaluation, reconsider and integrate the retrieved information with present information, and finally assign the formal evaluation. (De Nisi et al 1984). Differences between raters can occur at any one of these stages, and to ensure a level of similarity between the frontline, or junior managers one must address an outstanding range of moderators. A well designed questionnaire is only one of them.

Julia King, (1997) tells us “Project managers and executives at Claremont Technology Group Inc are getting plenty of **free** career advise these days” , she goes on to talk about an upward feedback system.

One of the factors to emerge from this research is that if an organisation wants an upward appraisal system that works well in terms of accuracy of appraisal and team building capabilities it will by no means be free!

Unless the organisation addresses such moderators as training, reward, evaluation apprehension (emphasis on development rather than evaluation), top management commitment, feedback, action planning on the basis of that feedback, information share, clarification of expectations, and involvement of all participants, the process will remain window dressing. All of these processes have budgetary considerations, but the cost of ignoring them is learning the hard way that X into Y will not go.

Word count: 1,692.

References:

Yammarino, F.J & Atwater, L.E (1997) Do managers see themselves as others see them? Organisational Dynamics. Spring. Pp35-44.

Mount, M.K. (1984) Psychometric properties of subordinate ratings of managerial performance. Personnel psychology Vol37. Pp 687-701.

London, M & Wohlers, A.J (1991) Agreement between subordinate and self-ratings in upward feedback. Personnel psychology. Pp375-389.

Foti, R (1990 August) The role of cognitive categories in supervisor versus self-ratings. Paper presented at the national meeting of the Academy of Management, San Fransisco.

Cook, M (1993). Personnel Selection & productivity. Chichester: John Wiley and Sons.

King, J (1997) Managers tweak style after they get staff feedback via notes. Computer world. P80.