
Rethinking The Performance Appraisal: Between Evaluational and Motivational Purposes

Bagus Riyono

Hofstra University, New York.

ABSTRACT

There are two main purposes of performance appraisal. First, it is done to evaluate the past performance of employees in order to give sufficient and necessary data for personnel decision making. This is called evaluational purpose. Second, it is meant to be one of the strategies to motivate employees in order to increase their productivity. This one is called motivational purpose (Roberts & Pavlak, 1996). Most of researches on performance appraisal focused on the appraisal processes that governed by the evaluational paradigm, while motivational purpose, which is also the ultimate purpose of performance appraisal (Roberts & Pavlak, 1996), is not widely researched yet (see appendix). This paper discusses the differences between those two paradigms and argue why the focus should be shifted from evaluational to motivational purpose. The proposed characteristics of a new performance appraisal design is also discussed.

Key Words : Evaluational Paradigm, Motivational Paradigm, Performance Targeting, Subjective Judgement, Comparison Standard.

Bagus Riyono was born in Surakarta, on June 12, 1963. He has paid a great study. After Working for ASTRA (to 1992) and continental Oil (to 1995), he returned to Gadjah Mada University, from where he graduated in 1991. He is now one of the junior faculty member and assigned to teach I/O Psychology, Management and Organisational as well as statistics. Bagus has also performed some research. Among them are Cultural Value Orientation and Job Satisfaction and Measurement for Leadership Quality. Currently, he is taking a graduate degree in I/O Psychology at Hofstra University, New York, under The Fulbright Grants.

EVALUATIONAL PURPOSE

Evaluational paradigm defines the performance appraisal as a process to assess employee's past performance and evaluate them based on certain criteria. The result of this process then to be used for personnel decision making such as, deciding on monetary rewards, transfers, assignments and layoffs or terminations (Dipboye, Smith, and Howell, 1994).

The kinds of problems that have to be addressed then are, what should be evaluated and how to evaluate them accurately. These lead to the issues of validity and reliability of a performance appraisal measures. Therefore, to develop a good perfor-

mance appraisal system, things that must be considered are, reliability, practicality, relevance, fairness and discriminativeness (Dipboye, Smith, and Howell, 1994).

PROBLEMS AND ATTEMPTS TO SOLVE

In order to develop a performance appraisal system that comply to those characteristics, researcher found what Dipboye, Smith and Howell (1994) call as *rating effects*, i.e. several systematic tendencies to rate in certain direction. These effects include halo, leniency, central tendency, severity, context, order, negativity and a lot more. Among those effects halo and leniency are the most widely considered rating effects in performance appraisal researches (Woehr and Huffcutt, 1994; Dipboye, Smith and Howell, 1994).

Halo effects is tendencies to give the same level of rating across all dimension. For example, "an employee who is seen as having a good attitude might be evaluated positively on not only attitude but other dimensions, such as quantity and quality of performance" (Dipboye, Smith and Howell, 1994).

Leniency effects are tendencies to give all employees toward the positive ratings. This could happen because the rater does not want to be considered as a mean person or, he/she personally likes the ratees. Villanova, Bernardin, Dahmus, and Sims (1993) call this situation as performance appraisal discomfort.

Beside those rating effects, other issue concerning evaluational purpose of performance appraisal is rating accuracy. Rating accuracy deals with the comparison of rater judgments against a standard or actual performance. For example if an employee has made all necessary effort to do his/her job, those efforts might not be considered by the rater because of not being informed or

having memory problems. That leads to inaccuracy in performance rating.

In order to manage and minimize these shortcomings in the process of performance appraisal, several rater training methods had been developed. Those methods are categorized in four kinds of rater training: rater error training, performance dimension training, frame-of-reference training and behavioral observation training (Woehr and Huffcutt, 1994).

Rater error training. Rater error training attempts to eliminate rating effects such as halo and leniency. The training guides the participants to recognize those rating effects and encourage to avoid them on the process of performance rating. This way it is believed that the rating process will be more effective and the rating effects will be minimized.

A meta analysis study by Woehr and Huffcutt (1994) showed that this type of training has failed to reduce rating effects effectively. The effect size (d) for reducing halo effects is .33 and leniency effects is only .21. Cohen (1977) suggests an effect size of .2 represents a small effect, an effect size of .5 represents a medium effect and effect size of .8 represents a large effect.

Performance dimension training. The performance dimension training deals with the dimension of performance that will be used in the ratings. The focus of the performance dimension training is the cognitive processing of information by the raters. Those cognitive processes are the key to rater training. The major premise is that an understanding of the way in which rater's process information with respect to evaluation will lead to training strategies that improve the effectiveness of performance ratings.

If raters trained recognize and use the appropriate dimensions on which rating will be required, this should lead to dimension

relevant judgment as opposed to a more global judgment. This process is believed to lead to more accurate rating across dimension. Thus the objective of this type of training is to increase accuracy in ratings.

The research results showed that this type of training does not effective enough to increase accuracy in rating. The effect size is even less than what is considered to be weak effect by Cohen (1977) with $d = .13$ (Woehr and Huffcutt, 1994).

Frame-of-reference training. The third kind of rater training is called frame-of-reference training. This training required a frame-of-reference as a standard to measure the accuracy of the rating. Trainee first being familiarized with certain standard of evaluation. Familiarity with these standards is believed to be able to lead to the increase of rating accuracy.

The frame-of-reference training is a more elaborate strategy than the performance dimension training. As an addition to performance dimensionality, it is also focus on performance standards. It is also a strategy of the social cognitive approach to performance appraisal.

In frame-of-reference training raters are trained with respect to common evaluation standards. The standards include a sample of behavioral incidents representing the dimension of performance.

This approach improves the results of performance dimension approach especially on the increase of rating accuracy. The effect size for accuracy increase is .83 (Woehr and Huffcutt, 1994). However, this approach does not improve other aspects as well (halo: $d = .13$; leniency: $d = .15$).

Behavioral observation training. The last approach to be discussed here called behavioral observation training. The idea is to eliminate or minimize the "on-line" effect of evaluation. In this approach rater is trained to differentiate between behavior observa-

tion and behavior evaluation. Observation process should be separated from evaluation process. On the observation process rater should only observe and take note on the ratee's behavior, or it is a data collection process. After data was collected, then the evaluational process will begin with analyzing the data and judging the quality and quantity of the behaviors related to performance.

The training process involve a memory measure. It also covers the ability to detect and the influence of perception to a specific behavioral events. In this training the measure is not on the rating accuracy but rather on the observational accuracy. The more accurate observation the more effective the evaluation to be, is the premise of this particular approach.

Research by Woehr and Huffcutt (1994) showed that the effect of the behavioral observation training to rating accuracy is between moderate to high effect ($d = .77$) but less effective in increasing observational accuracy ($d = .49$).

SHORTCUT IDEAS TO SOLVE THE PROBLEM

As attempts to recude those effects failed, some shortcut procedures was developed to force the fulfillment for one of the characteristics for a good appraisal governed by evaluational paradigm: distinctiveness. Those procedures are comparative procedures, which main purpose is to compare employees to each other on their job performance. Comparative procedures includes (a) ranking (b) paired comparison and (c) forced distribution (Dipboye, Smith and Howell, 1994).

Ranking puts employees in a list of quality performance. "They are ranked from the best to the worst on each dimension and/or on overall performance" (Dipboye,

Smith and Howell, 1994).

In paired comparison employees are being paired to all possible referent others. The appraiser make judgments on who is better in every pairs. Those who received more favorable judgments will be considered as having better performance.

Forced distribution groups employees as if their performance constitute a normal distribution. Based on the normal curve employees are groups in several performance categories (e.g. 1, 2, 3, 4, 5) with systematic number of quota in each group (e.g. 10% for 1, 20% for 2, 40% for 3, 20% for 4 and 10% for 5). In other words, only 10% of employees can receive the highest performance rating.

The main problem of comparative procedures is not only accuracy and fairness, but more on the culture it creates in the relationships among employees. In order to be the best, employees will compete against each other, while they should be cooperating. Deutsch (1949) in his experiment found that cooperation among colleagues showed characteristics that every organization would like to have:

1. Coordination of efforts.
2. Diversity in amount of contributions per member.
3. Sub-division of activities.
4. Achievement pressure.
5. Attentiveness to fellow members.
6. Mutual comprehension of communication.
7. Common appraisals of communication.
8. Orientation and orderliness.
9. Productivity per unit time.
10. Quality of products and of discussions.
11. Friendliness during discussion.
12. Favorable evaluation of the group and its products.

13. Group functions.
14. Perception of favorable effects upon fellow members.
15. Incorporation of the attitude of the generalized other.

It is obvious that the comparative procedures have been misleading. It is in conflict with the organization and employees goals. Deming (1992) said that "it will bring up the worst of people". Because if people cannot become better or improve themselves, then the other option to be number one is to make others worse than him/her. This way of thinking will lead to office politics and unnecessary conflicts that bring up the worst of people.

This mistake is due to a paradigm defect. It is not merely because of wrong processes or wrong techniques, but rather because of mistakenly focus on the wrong purpose of the process. These procedures force the evaluational purpose to be in place but neglect the motivational purpose, which is the main purpose of performance appraisal.

MOTIVATIONAL PURPOSE

The researches based on the evaluational paradigm mostly failed to provide evidences in order to improve the effectiveness of performance ratings. Halachmi (1993) stated that the focus on evaluational purpose in performance appraisal researches is "expensive, has limited value and may even be dysfunctional for improving future performance."

It is like a metaphor of trying to find a black cat in the dark room painted black, at night with no light. It requires a lot of energy and creativity, but only to find out that the cat is not in that room at all. The problem is not how to find the best way to do it, but we have to move to the other room. We must

change our focus to motivational purpose.

Instead of looking to the past, motivational purpose focus on the future. Halachmi (1993) proposes a concept of performance targeting as a replacement for the performance appraisal process. It is a process "which embraces a strategic perspective and an orientation toward the future." "Performance targeting shift the focus from documenting and evaluating an employee's work to assessing the partnership between a subordinate and a supervisor" (Halachmi, 1993).

Halachmi (1993) and Deming (1992) suggest to eliminate the performance appraisal system at all. They believe that it will bring out more harms than giving value to overall organizational goals and objectives.

However, if the performance appraisal system is eliminated at all, other problems may come up. How can we know the strength of the company's current human resources? What are promotion and job rotation based on? Looking to the past is also important as long as not being trapped by the past.

Therefore a new performance appraisal system is needed. A system that can learn from the past and manage the future of job performance. Agreement between employee and supervisor (rater) on the performance appraisal results is one of the essential criteria for this new system. If the employee feels that the rating is not fair he/she will be demotivated. This is well explained by Adam's equity theory. In motivational paradigm the consequences of the rating process or results is more important since it will lead to working harder or achieving higher productivity.

This process acknowledges subjective judgment. In the evaluational paradigm subjectivity is avoided because, "subjective judgments of performance tend to introduce a great deal of distortion into the measure-

ment process" (Woehr and Huffcutt, 1994). Since our purpose is not to evaluate, those distortion does not really matter as much. As long as both employee's and supervisor's combined subjectivities resulting in a higher motivation or drive to improve the productivity, subjectivity is accepted.

One value that can be derived from evaluational researches is that we can not eliminate subjective judgment. Therefore, we can not deny it. We have to accept subjectivity and work around it to fulfill our purpose.

Schrader and Steiner (1996) propose one strategy to achieve higher level of agreement between self and supervisory performance ratings. They argue that common comparison standards will lead to higher agreement. "A comparison standard represents the benchmark, or standard, against which a rater compares the ratee's performance" (Schrader and Steiner, 1996).

Lack of agreement between employee and supervisor on the performance rating is caused by the differential comparison standards. This occurs when "raters select different referent individuals, groups, or specific standards on which to base their rating." (Schrader and Steiner, 1996). Therefore, common comparison standards should be established in order to increase correlation between different raters ratings (see Schrader and Steiner, 1996).

There are five types of comparison standards, i.e. ambiguous, internal, absolute, relative, and multiple. Ambiguous comparison standards is applied the criteria stated in general terms and lack of operational definition. Internal comparison standards is comparison with self. Absolute comparison standards is comparison with some objective measure. Relative comparison standards is comparison with others in workgroup. Multiple comparison standards combine more than one standards.

Among these five, absolute and multiple comparison standards are the strongest in order to enable high correlation between self and supervisory ratings. With absolute comparison standards correlation between self and supervisory rating is $r = .50$, and with multiple comparison standard $r = .55$ (Schrader and Steiner, 1996).

CONCLUSION

As a conclusion, a new criteria for a good performance appraisal is proposed to be examined in future studies of performance appraisal. One big shift of focus should be taken. Instead of focusing on evaluational purpose, future study of performance appraisal should focus on motivational purpose, which investigates the influences of performance appraisal processes on motivation and higher productivity.

The new design or procedures of performance appraisal should eliminate comparison among employees, since the main purpose of it is to increase productivity. Increasing productivity cannot be achieved at Conclusion

As a conclusion, a new criteria for a good performance appraisal is proposed to be examined in future studies of performance appraisal. One big shift through internal competition. In order to do so, absolute or multiple comparison standard should be used instead of relative or internal standards.

Rater/supervisory training should focus on increasing skills for giving productive feedback and coaching to lead to the performance targeting strategy. .pa

REFERENCES

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Deming, W. E. (1992). *The 14 Points* [videorecording]. Chicago, Ill.: Films Incorporated.

Deutsch, M. (1949). An experimental study of the effects of cooperation and competition upon group process. *Human Relations*, 2, 199-232.

Dipboye, R. L., Smith, C. S. and Howell, W. C. (1994). *Understanding Industrial and Organizational Psychology: An Integrated Approach*. Fort Worth: Harcourt Brace College Publishers.

Halachmi, A. (1993). From performance appraisal to performance targeting. *Public Personnel Management*, 22, 323-344.

Roberts, G. E. and Pavlak, T. (1996). Municipal government personnel professionals and performance appraisal: is there a consensus on the characteristics of an effective appraisal system? *Public Personnel Management*, 25, 379-408.

Schrader, B. W. and Steiner, D. D. (1996). Common comparison standard: an approach to improving agreement between self and supervisory performance ratings. *Journal of Applied Psychology*, 81, 813-820.

Woehr, D. J. and Huffcutt, A. I. (1994). Rater training for performance appraisal: a quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189-205.

APPENDIX

- A. Reference samples on *evaluational* oriented performance appraisal researches, collected from Psychlit Database.
1. Balzer, W.K. and Sulsky, L.M. (1992). Halo and performance appraisal research: a critical examination. *Journal*

- of applied Psychology, 77, 975-985.
2. Conway, J.M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management*, 22, 139-162.
 3. Day, N.E. (1995). Can performance raters be more accurate? Investigating the benefits of prior knowledge of performance dimensions. *Journal of Managerial Issues*, 7, 323-342.
 4. DeNisi, A. S. and Petes, L.H. (1996). Organization of information in memory and the performance appraisal process: evidence from the field. *Journal of Applied Psychology*, 81, 717-737.
 5. Goffin, R.D. and Jackson, D.N. (1992). Analysis of multitrait-multirater performance appraisal data: composite direct product method versus confirmatory factor analysis. *Multivariate Behavioral Research*, 27, 363-385.
 6. Harris, H. M. (1994). Rater motivation in the performance appraisal context: a theoretical framework. *Journal of Management*, 20, 737-756.
 7. Hartel, C.E. (1993). Rating format research revisited: format effectiveness and acceptability depend on rater characteristics. *Journal of Applied Psychology*, 78, 212-217.
 8. Ilgen, D.R., Barnes-Farrell, J.L., and McKellin, D.B. (1993). Performance appraisal process research in the 1980s: what has it contributed to appraisals in use? *Organizational Behavior and Human Decision Processes*, 54, 321-368.
 9. Jackson, C. (1996). An individual differences approach to the halo-accuracy paradox. *Personality and Individual Differences*, 21, 947-957.
 10. MacLachlan, M. (1993). Depression is not necessary for pessimism in performance appraisal. *Medical Science Research*, 21, 171.
 11. Maroney, B.P. and Buckely, M.R. (1992). Does research in performance appraisal influence the practice of performance appraisal? Regrettably not. *Public Personnel Management*, 21, 185-196.
 12. Neck, C.P., Stewart, G.L. and Manz, C.C. (1995). Thought self-leadership as a framework for enhancing the performance of performance appraisers. *Journal of Applied Behavioral Science*, 31, 278-302.
 13. Ralston, R.W. and Waters, R.O. (1996). The impact of behavioral traits on performance appraisal. *Public Personnel Management*, 25, 409-421.
 14. Robert, G.E. and Pavlak, T. (1996). Municipal government personnel professionals and performance appraisal: is there a consensus on the characteristics of an effective appraisal system? *Public Personnel Management*, 25, 379-408.
 15. Sanchez, J.I. and De La Torre, P. (1996). A second look at the relationship between rating and behavioral accuracy in performance appraisal. *Journal of Applied Psychology*, 81, 3-10.
 16. Verma, A., DeNisi, A.S. and Peters, L.H. (1996). Interpersonal affect and performance appraisal: a field study. *Personnel Psychology*, 49, 341-360.
 17. Villanova, P., Bernardin, H. J., Dahmus, S. A. and Sims, R. L. (1993). Rater leniency and performance appraisal discomfort. *Educational and Psychological Measurement*, 53, 789-799.

18. Woehr, D. J. and Huffcutt, A. I. (1994). Rater training for performance appraisal: a quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189-205.
- B. Reference samples on motivational oriented performance appraisal researches, collected from Psychlit Database.
1. Hillman, L. W., Schwandt, D. R. and Bartz, D. E. (1990). Enhancing staff members' performance through feedback and coaching. *Journal of Management Development*, 9, 20-27.
2. Schrader, B. W. and Steiner, D. D. (1996). Common comparison standard: an approach to improving agreement between self and supervisory performance ratings. *Journal of Applied Psychology*, 81, 813-820.
3. Schay, B. W. (1993). In search of the Holy Grail: lessons in performance management. *Public Personnel Management*, 22, 649-668.
4. Waite, M. L., Newman, J. M. and Krzyzotofiak, F. J. (1994). Association among performance appraisal, compensation, and total quality programs. *Psychological Reports*, 75, 524-526.
5. Mullin, R. F. and Sherman, R. (1993). Creativity and performance appraisal: shall never the twain meet? *Creativity Research Journal*, 6, 425-434.
6. Halachmi, A. (1993). From performance appraisal to performance targeting. *Public Personnel Management*, 22, 323-344.

