

AN OVERVIEW OF PERFORMANCE APPRAISAL

USES OF PERFORMANCE APPRAISAL

Cleveland, J. N., Murphy, K. R., & Williams, R. E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology*, 74(1), 130-135.

- Information from PA had the greatest impact on salary administration, performance feedback, and the identification of employee strengths and weaknesses. Information from PA was used the least for evaluating personnel systems and as criteria for validation research.

PSYCHOMETRIC ISSUES

Murphy, K. R. and Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, CA: SAGE Publications, Inc.

- Accuracy is how much the ratee's score differs from a "true score" (the distance between the two). In PA, "true scores" are often given by subject matter experts (SME's). It is the biased-free score that a rater would give under optimal conditions

Murphy, K. R. and Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology*, 74(4), 619-624.

- The average correlation between error and accuracy was .05. Therefore, the authors do not recommend the use of rater error measures as indirect indicators of accuracy (which many researchers commonly do).

PERFORMANCE APPRAISAL INSTRUMENTS

Tziner, A. and Kopelman, A. (2002). Is there a preferred performance rating format?: A non-psychometric. *Applied Psychology: An International Review*, 51(3), 479-503.

- The psychometric advantages of BOS (behavior observation scale) over GRS (graphic rating scale) seem to be superior reliability and validity. BOS also appears to minimize communication barriers, role ambiguities, lead to clearer and more observable goals and consequent ratee commitment to carry them through. Raters are also more likely to feel comfortable with BOS because it directs attention to the frequency of specific and incontrovertible behaviors. The BARS (behaviorally anchored rating scales) was the least favorable of raters and made the least contributions to workplace improvements.

TRAINING THE RATER

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

- The authors evaluate four rater training strategies (rater error training, performance dimension training, frame-of-reference training, and behavior observation training) with respect to four dependent measures (halo, leniency, rating accuracy, and observational accuracy). Each of the four training strategies appears to be at least moderately effective and they increase rating accuracy. Overall, frame-of-reference (FOR) training tends to be the most effective. Combining different strategies can also be very effective.)

* Especially when it is combined w/ BOT.

RATER COGNITIVE PROCESSES

DeNisi, A. S. and Williams, K. J. (1988). Cognitive approaches to performance appraisal. *Research in Personnel and Human Resources Management*, 6, 109-155.

- The authors review the literature concerning the cognitive approach to studying PA. Cognitive approaches have resulted in researchers studying the subjective organization of information in memory, the distinction between on-line and memory-based ratings, the role of rater schemata, and the problems involved in rater's reprocessing of information collected for one purpose so that it can be used to make ratings.

RATER MOTIVATION

Harris, M. M. (1994). Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management*, 20(4), 737-756.

- There are situational variables (i.e., accountability, organizational HRM strategy, task/outcome dependence, trust, forms) and personal variables (i.e., amount of information, self-efficacy, mood) that affect motivational factors (i.e., rewards, negative consequences, impression management), which influence PA behaviors (i.e., observation, storage, retrieval, integration, rating, feedback).

Longenecker, C. O., Sims, H. P., & Gioia, D. A. (1987). Behind the mask: The politics of employee appraisal. *The Academy of Management Executive*, 1(3), 183-193.

- The most interesting finding from this study is that accuracy is not the primary concern of practicing executives in appraising subordinates. The main concern is how best to use the appraisal process to motivate and reward employees (i.e., getting them raises).

MODERATORS

Mero, N. P. and Motowidlo, S. J. (1995). Effects of rater accountability of the accuracy and favorability of performance ratings. *Journal of Applied Psychology*, 80(4), 517-524.

- Participants who were made to feel accountable by having to justify their ratings to the experimenter in writing rated their simulated subordinates more accurately. This finding is contradictory to what Dalessio (1998) argues (see below).

Jawahar, I. M. and Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, 50, 905-925.

- The results of this study clearly support the performance appraisal models (e.g. DeCotis & Petit, 1978; DeNisi, Cafferty, & Meglino, 1984; Taylor & Wherry, 1951) that predict that appraisal purpose influences appraisal leniency. PA ratings are nearly 1/3 of a standard deviation larger when ratings are obtained for administrative vs. research purposes. This is called the Performance Appraisal Purpose (PAP) effect.

Dobbins, G. H., Cardy, R. L., & Truxillo, D. M. (1988). The effects of purpose of appraisal and individual differences in stereotypes of women on sex differences in performance ratings: A laboratory and field study. *Journal of Applied Psychology*, 73(3), 551-558.

- This study focused on the social-cognitive processes (stereotypes) of raters to explore gender biases. The authors found that raters with traditional stereotypes of women evaluated female performances with a greater negative bias than raters with nontraditional stereotypes, but only when the ratings were for administrative decisions.

Saks, A. M. and Waldman, D. A. (1998). The relationship between age and job performance evaluations for entry-level professions. *Journal of Organizational Behavior*, 19, 409-419.

- Age was negatively related to job performance evaluations, but this relationship was completely eliminated after controlling for undergraduate grade average and the number of previously held jobs. This research suggests that other factors besides age bias might be affecting performance ratings.

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology*, 88(4), 694-706.

- In general, Whites received higher PA ratings than Blacks and Hispanics. Also, objective measures were associated with very similar, if not somewhat larger, standardized ethnic group differences than subjective measures across a variety of indicators, with Whites outperforming minority groups. This suggests a major societal concern.

360-DEGREE PERFORMANCE FEEDBACK

Dalessio, A. T. (1998). Using multisource feedback for employee development and personnel decisions. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 278-330), San Francisco: Jossey-Bass Publishers.

- There is more error when ratings are made for administrative purposes. Accuracy can be increased when raters are anonymous (rather than accountable).
- The ratees express more positive attitudes when feedback comes from multiple sources.
- Moderate performance improvements do result from multisource feedback processes.

Conway, J. and Huffcutt, A. I. (1997). Psychometric properties of multisource ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*(4), 331-360.

- The purpose of this study was to examine the psychometric properties of multisource ratings.
- General findings:
 1. Increasing the number of raters increases the reliability of ratings. Using the Spearman-Brown prophecy formula, the authors calculated that you can achieve a reliability of .70 if you use 6 subordinates, 3 supervisors, or 4 peers.
 2. Composite ratings (as opposed to overall or dimensional ratings) were found to maximize reliability.
 3. Subordinates showed the lowest reliability in their ratings and the lowest correlations with other sources. However, they do provide a unique perspective because their ratings are not redundant with other sources.
 4. Self-ratings showed the lowest correlations with other sources. This may be because of egocentric biases (inflated ratings) or self-esteem effects.
 5. Job type was found to be a moderator. There were clear mean differences between managerial and non-managerial jobs. Managerial jobs displayed lower correlations because these jobs are more complex and harder to rate.

360-Degree Feedback Results

Mean Correlations	Self-Peer	Self-Supervisor	Peer-Supervisor
Harris & Schaubroeck (1988)*	.24	.22	.48
Conway & Huffcutt (1997)	.19	.22	.34

*Harris, M. M. and Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.

PERCEIVED FAIRNESS OF PERFORMANCE APPRAISAL

Bartol, K. M., Durham, C. C., & Poon, J. M. L. (2001). Influence of performance evaluation rating segmentation on motivation and fairness perceptions. *Journal of Applied Psychology, 86*(6), 1106-1119.

- Explored the impact of rating segmentation on motivation and perceptions of fairness. They had two conditions: (1) low segmentation—three categories (1=below average, 2=average, 3=above average) and (2) moderate segmentation—five categories (1=comparatively low, 2=below average, 3=average, 4=above average, 5=exceptionally high).
- The 5-category system resulted in higher self-efficacy regarding their ability to reach the next higher rating category, higher goals for rating improvement, and higher rating improvements than the 3-category system.
- Fairness (distributive justice) perceptions at the lower range in the 5-category system were significantly lower than for the counterpart lower range individuals in the 3-category system.

Greenberg, J. (1986). Determinants of perceived fairness of performance evaluations. *Journal of Applied Psychology*, 71(2), 340-342.

- This is one of the most widely cited articles in the performance appraisal literature.
- The overall message of this study is that we need to look at procedural justice (in addition to distributive justice).

*Creating Fairness in Performance Management Systems**

Type Of Justice	Determinant of Fairness	Performance Management Process		
		System Development	Appraisal Processes	Feedback Delivery
Procedural	Voice	+	+	+
	Consistency	+	+	
	Challenge/Appeal			+
	Familiarity		+	
	Communication of Expectat.	+	+	
Interactional	Respect			+
	Feedback		+	+
	Timeliness			+
Distributive	Relevance	+	+	+
	Bias		+	
	Outcome		+	

*The information in this chart was taken from Greenberg (1986) and:

Gilliland, S. W. and Langdon, J. C. (1998). Creating performance management systems that promote perceptions of fairness. In J. W. Smither (Ed.), *Performance appraisal: State of the art in practice* (pp. 209-243), San Francisco: Jossey-Bass Publishers.

- This chart suggests that by just giving a ratee “voice” in the PA processes, fairness perceptions can be improved.

Performance Appraisal Readings

This is a nice, short one that everyone could read quickly:

Cleveland, J.N., Murphy, K.R., & Williams, R.E. (1989). Multiple uses of performance appraisal: Prevalence and correlates. *Journal of Applied Psychology, 74*(1), 130-135.

I tried to capture all of the different (broad) issues concerning PA by picking the most comprehensive article from each week on our syllabus:

Murphy, K.R. and Balzer, W.K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology, 74*(4), 619-624.

Tziner, A. and Kopelman, A. (2002). Is there a preferred performance rating format?: A non-psychometric. *Applied Psychology: An International Review, 51*(3), 479-503.

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

DeNisi, A.S. and Williams, K.J. (1988). Cognitive approaches to performance appraisal. *Research in Personnel and Human Resources Management, 6*, 109-155.

Harris, M. M. (1994). Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management, 20*(4), 737-756.

Jawahar, I. M. and Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology, 50*, 905-925.

Conway, J. and Huffcutt, A. I. (1997). Psychometric properties of multisource ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*(4), 331-360.

Bartol, K. M., Durham, C. C., & Poon, J. M. L. (2001). Influence of performance evaluation rating segmentation on motivation and fairness perceptions. *Journal of Applied Psychology, 86*(6), 1106-1119.