Mishra, V., & Roch, S. G. (2013). Cultural values and performance appraisal: Assessing the effects of rater self-construal on performance ratings, *The Journal of Psychology, 147*, 325–344.

Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *Journal of Applied Psychology, 89*, 158–164.

Ng, K., Koh, C., Ang, S., Kennedy, J. C., & Chan, K. (2011). Rating leniency and halo in multisource feedback ratings: Testing cultural assumptions of power distance and individualism-collectivism. *Journal of Applied Psychology, 96*, 1033–1044.

Peretz, H., & Fried, Y. (2012). National cultures, performance appraisal practices, and organizational absenteeism and turnover: A study across 21 countries. *Journal of Applied Psychology, 97*, 448–459.

Ployhart, E., Wiechmann, D., Schmitt, N., Sacco, J. M., & Rogg, K. (2003). The cross-cultural equivalence of job performance ratings. *Human Performance, 16*, 49–79.

Shipper, F., Hoffman, R. C., & Rotondo, D. M. (2007). Does the 360 feedback process create actionable knowledge equally across cultures? *Academy of Management Learning & Education, 6*, 33–50.

Suh, E. M., Diener, E., & Updegraff, J. A. (2008). From culture to priming conditions: Self-construal influences on life satisfaction judgments. *Journal of Cross-Cultural Psychology, 39*, 3–15.

Toegel, G., & Conger, J. A. (2003). 360-degree assessment: Time for reinvention. *Academy of Management Learning & Education, 2*, 297–311.

Tsui, A. S. (2006). Contextualization in Chinese management research. *Management and Organization Review, 2*, 1–13.

Varma, A., Pichler, S., & Srinivas, E. S. (2005). The role of interpersonal affect in performance appraisal: Evidence from two samples—U.S. and India. *International Journal of Human Resource Management, 16*, 2029–2044.

Zhou, J., & Martocchio, J. J. (2001). Chinese and American managers' compensation award decisions: A policy-capturing comparative study. *Personnel Psychology, 54*, 115–145.

# Disappointing Interventions and Weak Criteria: Carving Out a Solution Is Still Possible

Aharon Tziner
*Netanya Academic College*

Sylvia G. Roch
*University at Albany, State University*

In their focal article, Adler and his colleagues (2016) elaborate on the pros and cons of abolishing the performance appraisal process in organizations.

Aharon Tziner, Schools of Behavioral Studies and Business Administration, Netanya Academic College, Netanya, Israel; Sylvia G. Roch, Department of Psychology, University at Albany, State University.

Correspondence concerning this article should be addressed to Aharon Tziner, Schools of Behavioral Studies and Business Administration, Netanya Academic College, 1 University Street, Netanya, Israel. E-mail: atziner@netanya.ac.il

Sherman-Garr (2014) contends that this trend is on the rise because both managers—the raters—and their subordinates—the ratees—disdain performance scores. Employees feel that performance ratings do not reflect their actual performance, and therefore they do not gain the rewards they merit. Conversely, their supervisors/managers experience a great deal of frustration because the improvement of employee performance does not match up to the excessive time and effort invested in the appraisal process, making the whole process ineffective and inefficient. We agree that performance appraisals, specifically the practice of assigning performance ratings, appear to be a disliked and ineffective human resource function. However, we do not agree that goal attainment should be used in place of performance ratings; rating format and rater training represent "disappointing interventions" and, most of all, only "weak" criteria exist for performance ratings.

### Goal Attainment as an Alternative to Performance Ratings

One favored solution appears to be assessing an employee's performance according to his/her completion of assigned work goals in comparison with other employees, while accounting for the complexity and impact of these goals to the overall performance of the organization. In this case, the employee's compensation is linked to the degree of goal attainment adjusted, as noted, according to the complexity and impact on the organization's overall performance. Thus, performance achievement is determined by a global measure of goal attainment, namely by **end results.** Is this indicative of true performance? Almost 30 years ago, Landy and Farr (1983) pointed to a striking finding: Many "objective" measures exhibit low levels of reliability and consistency across equivalent indices. Interestingly, the correlation between different indices of absenteeism nears zero. Moreover, objective measures of output, sales, and the like are not available for each job, especially for managerial jobs. Furthermore, when they are available, they usually exhibit criterion deficiency because some components of work performance do not lend themselves to measurement, whereas others are contaminated by contextual factors.

For instance, we suspect that sales of detection devices of explosive materials will grow exponentially in world regions susceptible to be struck by terror acts. Thus, the volume of sales will hike, leading to an assessment that the performance of the salespeople has been exceptional, whereas in fact the performance level should be attributed to circumstantial factors. Would it be accurate to ascribe this sort of performance excellence solely to the competence and efforts of the salespeople?

Conversely, if an employee has done his/her best, yet nonetheless has failed to attain his/her set goals due to uncontrollable factors, would it be just and fair to penalize him/her? Ignoring deficiencies in the process leading

to goal attainment and the embedded context would hamper performance improvement and accurate feedback. Only if we center on the **process** rather than on end results can we detect what has gone wrong, provide punctual and credible feedback, and thereby hope to enhance employee performance. Thus, it is no surprise that performance appraisal experts for the past 20+ years have made a strong case for concentrating on behaviors versus results (e.g., Murphy & Cleveland, 1995).

### Two "Disappointing Interventions"

We contend that the two interventions discussed by Colquitt, Murphy, and Ollander-Krane as "disappointing," rating format and rating training, are anything but disappointing. It is true, unfortunately, that empirical findings have not corroborated the psychometric superiority of one rating format over others. So, is this a justifiable reason to label rating format research as disappointing? Tziner and his colleagues (Tziner, Joanis, & Murphy, 2000; Tziner, Kopelman, & Livneh, 1993; Tziner & Latham, 1989; Tziner & Murphy, 1999) have shown that too much weight has been placed on rating accuracy. Over 30 years ago, Bernardin and Beatty (1984) pointed out that ratees' reactions to appraisal systems are more likely than their psychometric qualities to make a significant contribution to sustaining the viability of appraisal systems. Regardless of the accuracy and psychometric characteristics of the performance ratings, an appraisal system will be rendered useless, and probably sink into decay, if it does not elicit **positive reactions** from both raters and ratees (Hedge & Borman, 1995).

Thus, it is not surprising that as Levy and Williams (2004) state in their review "Perhaps no area within the PA (performance appraisal) literature has seen such a dramatic increase in research attention since 1990 as ratee reactions to PA processes" (p. 889). Specifically in regard to rating formats, Tziner and colleagues found in a series of articles that the type of rating format used can influence goal clarity, goal acceptance, and goal commitment satisfaction (e.g., Tziner, Kopelman, & Joanis, 1997; Tziner et al., 1993; Tziner & Latham, 1989). Roch, Sternburgh, and Caputo (2007) found that, in general, absolute formats (which compare individuals with standards) are seen as more fair than relative formats (which compare individuals with their peers) and that format differences can influence perceived interpersonal justice, especially if employees do not trust their supervisors (Roch, 2015). Thus, it appears that the type of rating format used can influence employees' attitudes and justice perceptions, both of which have implications for employee performance, both task performance and organizational citizenship behaviors. Improved job satisfaction, organizational commitment, and so forth are worth real dollars (Cascio, 2000).

Even though, as mentioned by Colquitt and colleagues, we cannot determine in an organizational setting whether one rating format can more accurately assess employee performance than another format, we can evaluate whether a rating format is useful in promoting positive organizational attitudes and behavior. Type of rating format does matter.

Rater training also does matter, especially frame-of-reference (FOR) training. We were surprised that Colquitt and colleagues discussed rater error training (RET) in depth. Woehr and Huffcutt (1994) pointed out the problems with RET over 20 years ago, and as mentioned in the rater training meta-analysis by Roch, Woehr, Mishra, and Kieszczynska (2012), RET training has practically disappeared from the literature in the last 20 years.

It has been well established that FOR training improves rating accuracy (Roch et al., 2012; Woehr & Huffcutt, 1994). However, we believe that the main advantage of FOR training is that it helps to bring everyone "on the same page"—gives everyone, both raters and ratees, the same definition of performance. It is not important whether the definition is the "accurate" one, just that the definition is the one espoused by the organization. Raters may still see different aspects of employee performance and thus disagree, but at least they will have a common definition of performance. It is no surprise that in the last 10 years, FOR has been implemented in a wide variety of contexts, including assessing language proficiency (Dierdorff, Surface, & Brown, 2010), modeling competency (Lievens & Sanchez, 2007), and evaluating biodata items (Lundstrom, 2007).

Rater training may also help with the underlying problem of performance appraisal systems: rater motivation. We have long known that performance ratings are viewed as a management tool, one used to achieve certain ends, and not a measurement tool (Longenecker, Sims, & Gioia, 1987). Research has shown that performance rating inaccuracy is linked not only to rating format deficiencies and raters' cognitive impairment but also to raters' deliberate, volitional distortion of performance ratings. These distortions emanate from a gamut of motives and considerations, mostly related to a wish to promote valuable individual goals (e.g., supervisors avoid giving performance ratings that may antagonize employees; supervisors give low performance ratings because they fear that their employees will be transferred to another boss; supervisors avoid giving low performance ratings because they fear violent behavior on the part of their employees).

Rater training cannot deter raters from providing distorted ratings but can give raters the tools needed to provide ratings consistent with the organization's viewpoints and, more important, convey a message that the organization values performance ratings enough to invest in rater training. In a recent survey of 101 U.S. firms, Gorman, Meriac, Ray, and Roddy (2015) found that over 76% of these firms used rater training, with FOR training as

the most common type of training. Even more important, human resource executives whose firms provided rater training rated their performance appraisal systems as more effective than those not offering training. Controlling for firm size, the firms offering rater training had higher revenue than firms not offering rater training. Colquitt, Murphy, and Ollander-Krane remark that "However, neither variation on rater training has been successful in markedly improving ratings in organizations" (Adler et al., p. 225). However, data collected 20 years later present a different picture. Even though we do not know whether rater training has improved performance ratings in organizations, it appears that rater training can improve perceived performance appraisal effectiveness.

### "Weak Criteria"

Colquitt, Murphy, and Ollander-Krane suggest that two criteria can be used to evaluate performance ratings: rating accuracy and rating agreement. However, these are the proverbial "strawmen," easily knocked down. Murphy, Balzer, Sulsky, and colleagues wrote an excellent series of articles over 20 years ago suggesting that outside of laboratory contexts, accuracy and rater errors should not be used to evaluate performance ratings (e.g., Balzer & Sulsky, 1992; Murphy & Balzer, 1989; Murphy, Jako, & Anhalt, 1993; Sulsky, & Balzer, 1988). There is almost no disagreement regarding this assessment; why are Colquitt and colleagues making this point again today?

Similarly, there is a growing acceptance of the ecological validity argument in regard to performance ratings (e.g., Lance, Baranik, Lau, & Scharlau, 2009). Proponents of this argument contend that given individual raters' differences in respect to expectations, goals, and so on, raters may be more or less attuned to specific behaviors. Thus, lack of rater agreement may be a result of differences in the raters' perspectives and represent unique, but valid, observations rather than error. To a certain extent, rater differences in respect to expectations, goals, and personal characteristics can be improved by means of training, such as training to enhance one's self-efficacy, but raters will still see different behaviors (not all raters will be watching ratee performance at the same time, and individual performance varies). Thus, rating agreement is not a useful criterion for evaluating the quality of performance ratings. Given the problems with both rating accuracy and rater agreement, it is no surprise that, as mentioned earlier, research investigating ratee reactions to performance appraisal processes has seen such growth since 1990 (Levy & Williams, 2004).

Organizational interventions should be designed and conducted to change the organization and performance appraisal systems' characteristics, thereby leading to an increase in the degree of positive perceptions toward these entities. Such changes will subsequently affect rating behavior.

Our contention is this: Accuracy is not the most important thing. We, in organizational behavior and human resource management, put enormous efforts into the attempt to find the practices through which we can improve work behaviors and attitudes. If so, why not judge the value of interventions such as rating format and rater training in terms of their ability to improve work attitudes and behaviors, improvement that has financial value? So, let us not throw out performance ratings because of their perceived inaccuracy (which may or may not reflect reality), but let us judge it by its value in improving work behaviors and attitudes, which consequently improve the organization's performance. This, as mentioned, has financial value.

## References

Adler, S., Campion, M., Colquitt, A., Grubb, A., Murphy, K., Ollander-Krane, R., & Pulakos, E. D. (2016). Getting rid of performance ratings: Genius or folly? A debate. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 9*(2), 219–252.

Balzer, W. K., & Sulsky, L. M. (1992). Halo and performance appraisal research: A critical examination. *Journal of Applied Psychology, 77*, 975–985.

Bernardin, H. J., & Beatty, R. W. (1984). *Performance appraisal: Assessing human behavior at work*. Boston, MA: Kent.

Cascio, W. (2000). *Costing human resources: The financial impact of behavior in organizations* (4th ed.). Cincinnati, OH: Southwestern.

Dierdorff, E. C., Surface, E. A., & Brown, K. G. (2010). Frame-of-reference training effectiveness: Effects of goal orientation and self-efficacy on affective, cognitive, skill-based, and transfer outcomes. *Journal of Applied Psychology, 95*, 1181–1191.

Gorman, C. A., Meriac, J. P., Ray, J. L., & Roddy, T. W. (2015). Current trends in rater training: A survey of rater training programs in American organizations. In B. J. O'Leary, B. L. Weathington, C. J. L. Cunningham, & M. D. Biderman (Eds.), *Trends in training* (pp. 1–23). Newcastle upon Tyne, UK: Cambridge Scholars.

Hedge, J. W., & Borman, W. C. (1995). Changing conceptions and practices in performance appraisal. In A. Howard (Ed.), *The changing nature of work* (pp. 451–481). San Francisco CA: Jossey-Bass.

Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E. A. (2009). If it ain't trait it must be method: (Mis)application of the multitrait–multimethod methodology in organizational research. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Received doctrine, verity, and fable in organizational and social research* (pp. 339–362). New York, NY: Routledge.

Landy, F. J, & Farr, J. L. (1983). *The measurement of work performance*. New York, NY: Academic Press.

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management, 30*, 881–905.

Lievens, F., & Sanchez, J. I. (2007). Can training improve the quality of inferences made by raters in competency modeling? A quasi-experiment. *Journal of Applied Psychology, 92*, 812–819.

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

Lundstrom, J. T. (2007). *A new use of frame-of-reference training: Improving reviewers' inferences from biodata information* (Unpublished doctoral dissertation). Kansas State University, Manhattan, KS.

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

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

Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology, 78*, 218–225.

Roch, S. G. (2015, April). Rating format and justice: Importance of justice dimension and trust. In C. A. Gorman. (Chair), *It's okay to do rating format research again*. Symposium conducted at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Roch, S. G., Sternburgh, A., & Caputo, P. (2007). Absolute versus relative performance rating formats: Implications for fairness and organizational justice. *International Journal of Selection and Assessment, 15*, 302–316.

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology, 85*, 370–395.

Sherman-Garr, S. (2014). *Abolishing performance scores: A practical guide*. Oakland, CA: Bersin by Deloitte.

Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology, 73*, 497–506.

Tziner, A., Joanis, C., & Murphy, K. (2000). A comparison of three methods of performance appraisal with regard to goal properties, goal perception, and ratee satisfaction. *Group and Organizational Management, 25*, 175–190.

Tziner, A., Kopelman, R. E., & Joanis, C. (1997). Investigation of raters' and ratees' reactions to three methods of performance appraisal: BOS, BARS, and GRS. *Canadian Journal of Administrative Sciences, 14*, 396–404.

Tziner, A., Kopelman, R. E., & Livneh, N. (1993). Effects of performance appraisal format on perceived goal characteristics, appraisal process satisfaction, and changes in rated job performance: A field experiment. *Journal of Psychology, 127*, 281–292.

Tziner, A., & Latham, G. P. (1989). The effects of appraisal instrument, feedback, and goal-setting on worker satisfaction and commitment. *Journal of Organizational Behavior, 10*, 145–153.

Tziner, A., & Murphy, K. (1999). Additional evidence of attitudinal influences in performance appraisal. *Journal of Business and Psychology, 13*, 407–420.

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