2024

Decoding Variance and Predictive Ability in Selection Systems: An Application of Gauthier's Framework of Rater Cognitions

Peyton Anderson
Illinois State University

Decoding variance and predictive ability in selection systems: An application of Gauthier's framework of rater cognitions

Peyton Anderson

Department of Psychology, Illinois State University, Normal, IL, USA
Email: pmande4@ilstu.edu

Foster et al. (2024) offer compelling insights into the advancement of selection measurement assessments, emphasizing high construct validity and accurate predictive capabilities for evaluating candidates' work performance. Previous arguments surrounding the validity of selection assessments and the predictive performance abilities they encompass have been centered on viewing the measurement system as the culprit of error or have placed a heavy emphasis on understanding the noncognitive influences of raters, such as social contexts (Spence & Keeping, 2011). Foster et al. (2024) provide six recommendations of ways in which the field may benefit from shifting focus away from the measurement tool but rather toward other factors. Instead of elaborating on all recommendations provided in the focal article, this commentary will address and elaborate upon the broadly mentioned first recommendation and propose an additional method for understanding the root causes of rater error and variance in ratings. This commentary concludes with how to make approaches to predicting variance and correcting for error more specific to the cognitive processes of the rater.

Understanding rater cognitions

Foster et al. (2024) mention the importance of understanding the percentage of variance in ratings due to ratee main effects (dealing with the person being rated). However, the authors should have noticed a potential resource for understanding the variance involving examining rater cognition. The authors state that a large amount of the variance in performance ratings stems from their lack of being ratee related or a lack of variance due to ratee main effects. Specifically, Foster et al. (2024) reported that Scullen et al. (2000) found that 21%–25% of the variance in performance ratings was due to ratee main effects. This finding reiterates the importance of controlling for bias and rater idiosyncrasy when using predictive performance measures to make selection decisions or predictions about candidate performance for the role in question.

Previous literature has focused on how variance in ratings can be due to rater error, referring to distortion of ratings either consciously or subconsciously via politics, impression management, leniency error, and motivational influences (Spence & Keeping, 2011). Many different approaches to understanding how rater cognition or judgment processes can account for variance in performance ratings have been proposed in research (e.g., Gingerich et al., 2011; Lewis, 2021; Spence & Keeping, 2011). Gauthier et al. (2016) introduced a general approach involving understanding how rater cognition occurs in a three-phase process. The first phase involves the rater generating impressions about the ratee and formulating their inferences based on the

different dimensions of the competencies being assessed (i.e., the observation stage). During the second phase (i.e., the processing phase), the rater relies on schemas based on their concept of the assessed competence to categorize behavior. Last, in the integration phase, the rater will weigh the information gathered and translate their judgments into assessment scales (Lewis, 2021). It is necessary to understand that rater cognitions and individual differences influence the raters' behavior and cognitive processes during the three stages mentioned by Gauthier et al. (2016). Utilizing Gauthier et al. (2016) three-stage model of rater cognition processes would contribute to the issue introduced by Foster et al. (2024) Additionally, understanding the differences in rater interpretations of performance could indicate if those differences represent errors in judgment or offer meaningful insight into the variability of ratings (Sebok & Syer, 2015).

Rater training programs

Much attention focuses on rater training as a primary strategy for mitigating errors in performance judgments. However, DeNisi and Murphy (2017) provided an essential finding that when rater training programs aimed to alter rater perspectives to mitigate error, the programs were deemed ineffective. An alternative approach for eliminating error should be centered around training raters to hold consistent conceptions of good and poor performance, which produced beneficial results, according to DeNisi and Murphy (2017). Consistent with this approach is the notion that raters develop different schemas surrounding the competencies being assessed in their performance, according to the rater cognition that occurs within the observation and processing phase of Gauthier et al. (2016) framework (Lewis, 2021). Therefore, not only should rater training involve maintaining consistent perceptions of good and poor performance among the raters, but an additional focus on maintaining consistent perceptions of the assessed competencies should be considered.

Considering the type of processing raters engage in while providing performance ratings and implementing features of rater training programs that specifically target those processes is a practical step toward reducing bias and rater error that greatly saturate performance ratings. For example, Mills (1998) investigated the effects of frame-of-reference training (FOR) on the raters' ability to recall behavioral incidents and provide accurate performance ratings. The results indicated that raters who received the FOR training provided ratings with higher behavioral accuracy and recalled a significantly greater number of behaviors than subjects in the no-training condition (Mills, 1998, p. 79). Utilizing training that focuses on enhancing the cognitive processing of raters might be an effective strategy to gain more accurate performance ratings and control for high variance due to rater main effects. Nevertheless, efforts to improve the accuracy of rater processing and rater ability will not be without merit and will likely result in enhanced employee selection systems. Further research is warranted to investigate the effectiveness of various rater training programs to determine the optimal strategy for managing variations in the performance ratings of candidates.

Methods for predicting and correcting rater error

In later research, Gingerich et al. (2018) identified "contrast effects" as an established source of variability in assessment judgments. Contrast effects occur when scores deviate from the portrayed level of competency in a previous interaction (Gingerich et al., 2018). According to Gingerich et al. (2018), the shift could result from the range frequency of raters' internal scales or the emphasized performance aspects within assessment judgments. Therefore, understanding raters' internal scales of performance aspects may indicate if biased or inaccurate ratings are likely to be present and to establish the optimal rater training procedures for reducing the variance. Interestingly, Gingerich et al. (2018) propose exposing raters to reference performances immediately before assessment as a solution.

A potential strategy to understand the assessment judgments of raters could be to gather information capturing the raters' cultural self-construal. According to Mishra and Roch (2013), measuring the effects of cultural values on performance ratings can be useful for understanding the schemas or preferences that raters form surrounding the competencies of interest in the performance evaluation. Furthermore, Mishra and Roch's (2013) findings indicate that measuring a rater's self-construal (i.e., independent vs. interdependent) within their culture will provide further predictive ability on the evaluations they form. The results of the study indicated that the raters with a strong interdependent or collectivist self-construal tend to provide more positive ratings for ratees who also display interdependent traits (Mishra & Roch, 2013). Investigating the impact of self-construal on perfromance ratings can offer valuable insights into understanding the cognitive processes of raters.

An additional idea could be to administer self-reported measures of ideas about the assessed competencies in the job of interest to the raters before analyzing their performance ratings. Interpreting the rater's reported perspectives would allow for a better understanding of the variance that occurs among performance ratings, better analysis and implementation of the ratings, and enhanced organizational knowledge about correcting errors in their selection system. To better understand rater variance, Sebok and Syer (2015) investigated rater interpretations among different noncognitive attributes in performance assessments. As a result, Sebok and Syer (2015) concluded that the substantial variance in ratings may result from raters' possession of different interpretations of performance. Examining the personality dimensions of each rater is one strategy to provide an explanation for the variance in rater interpretations of performance. Findings from a meta-analysis on rater personality traits indicated that rater personality accounted for between 6% to 22% of the variance in perfromance ratings (Harari et al., 2014b). However, research on gathering rater interpretations before undergoing the performance appraisal process has remained limited; it is essential to understand its impact on ratings comprehensively.

A method of anticipating rater variance and correcting rater error that has yet to be examined could be to incorporate Gauthier et al. (2016) framework when analyzing the variance in performance ratings. Individual differences in raters are meaningful and can explain much about the root source of error in rater judgments (Gingerich et al., 2018). By breaking down the three phases of Gauthier et al. (2016) framework and understanding the individual differences of raters that contribute to variances in rater cognition among each phase, test developers will become aware of which stage of the rating process to focus their error mitigation efforts. For example, suppose it is established that during the processing stage, a standard schema the raters rely on is the idea that extroversion is predictive of effective communication abilities. In that case, we can anticipate that during the processing phase, the raters will likely rate highly extroverted employees as high in performance when communication is the competency of interest. Gathering information about rater schemas and predispositions regarding the assessed competencies will likely require further research.

Concluding thoughts

Foster et al. (2024) begin a meaningful discussion with their first recommendation. However, understanding how rater cognition patterns are formed should additionally be considered. Identifying patterns of rater cognitions can provide a beneficial instrument to address common factors influencing ratings (Lewis, 2021). Implementing Gauthier et al. (2016) three-phase framework can allow for a better answer on how rater idiosyncrasy influences ratings through patterns of rater cognition. Once the common mechanisms or characteristics of raters become evident within each pattern of rater cognition, organizations may gain the predictive ability to know when rater error is likely to occur in their selection systems and how significant the variance

in performance ratings will likely be. With this predictive knowledge, selection systems can work more efficiently by correcting for the appropriate type of rater error rather than relying on rater training to mitigate error.

References

DeNisi, A. S., & Murphy, K. R. (2017). Performance appraisal and performance management: 100 years of progress? *Journal of Applied Psychology*, **102**(3), 421–433. https://doi.org/10.1037/apl0000085

Foster, J., Steel, P., Harms, P., O'Neill, T., & Wood, D. (2024). Selection tests work better than we think they do and have for years. *Industrial and Organizational Psychology Journal*, **17**(3), 269–282.

Gauthier, G., St-Onge, C., & Tavares, W. (2016). Rater cognition: review and integration of research findings. *Medical Education*, **50**(5), 511–522. https://doi.org/10.1111/medu.12973

Gingerich, A., Regehr, G., & Eva, K. W. (2011). Rater-based assessments as social judgments: Rethinking the etiology of rater errors. *Academic Medicine*, **86**, 1–7. https://doi.org/10.1097/acm.0b013e31822a6cf8.

Gingerich, A., Schokking, E., & Yeates, P. (2018). Comparatively salient: Examining the influence of preceding performances on assessors' focus and interpretations in written assessment comments. *Advances in Health Sciences Education*, **23**(5), 937–959. https://doi.org/10.1007/s10459-018-9841-2

Harari, M. B., Rudolph, C. W., & Laginess, A. J. (2014b). Does rater personality matter? A meta-analysis of rater big five-performance rating relationships. *Journal of Occupational and Organizational Psychology*, **88**(2), 387–414. https://doi.org/10.1111/joop.12086

Lewis, H. A. (2021). *A study to identify factors of rater cognition from different rater perspectives during performance-based assessments*. University of Kansas ProQuest Dissertations Publishing.

Mills, N. E. (1998). *A cognitive approach to rater training*. Kansas State University.

Mishra, V., & Roch, S. G. (2013). Cultural values and performance appraisal: Assessing the effects of rater self-construal on performance ratings. *Journal of Psychology*, **147**(4), 325–344. https://doi.org/10.1080/00223980.2012.694377

Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, **85**, 956–970.

Sebok, S. S., & Syer, M. D. (2015). Seeing things differently or seeing different things? Exploring raters' associations of noncognitive attributes. *Academic Medicine*, **90**, 1–6. https://doi.org/10.1097/acm.0000000000000902.

Spence, J. R., & Keeping, L. (2011). Conscious rating distortion in performance appraisal: A review, commentary, and proposed framework for research. *Human Resource Management Review*, **21**(2), 85–95. https://doi.org/10.1016/j.hrmr.2010.09.013