CAMBRIDGE
UNIVERSITY PRESS

COMMENTARY

Do selection tests "really" work better than we think they do?

S. Burak Ozkum

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

Foster et al. (2024) pointed out an alternative way to interpret the predictive power of selection tests, a critical topic that has kept our field busy for a long time. In their paper about the validity of the selection tests, the authors emphasized the importance of using variance components directly relevant to the ratees in job performance ratings instead of using overall ratings that might be contaminated. Criterion contamination is a frequently discussed issue in our field, yet we continue relying on noisy criterion variables, such as supervisory ratings of job performance. As discussed in the focal article, the scientific evidence suggests that ratee main effects account for roughly one-fourth of the variance in performance ratings, leaving a solid 75% variance explained by other factors, including rater effects, rater by ratee interactions, and residual variance.

Foster et al. (2024) pointed out the bright side of this issue by stating that the selection tests we have been using might have worked better than we thought. They enthusiastically suggested that when ratee-related variance is used as the criterion to test the predictive power of the eight commonly used selection tests, such as cognitive ability tests, personality inventories, emotional intelligence assessments, and structured interviews, the variance explained in the criterion exceeds the most optimistic estimations. At first glance, this is excellent news, considering how impactful the relatively lower validity coefficients reported by Sackett et al. (2022) were in our field, testing our faith in the validity of selection tests. The focal article's authors announced that commonly used predictors could explain 66% of the relevant variance in job performance ratings without correcting the validity coefficient for unreliability and range restriction. Before starting to celebrate our success as a field, we should carefully analyze the logic followed by the authors to understand how they came up with this conclusion.

The validity coefficient matrix used by the authors is provided in Table 1. They performed an ordinary least squares (OLS) multiple regression using these coefficients, yielding an R^2 of .164. Using the 25% ceiling of ratee-relevant variance, they concluded that this explained variance equals 66% of the variance in job ratings. Using the same logic and assuming that the variability in job performance ratings comes from the four primary sources identified by Jackson et al. (2020): ratee-related variance (29.77%), rater-related variance (30.07%), rater–ratee interaction (12.18%), and other variance (27.98%), the eight predictors used by the authors should explain 55.09% of the variability in the latent variable of job performance. Such conclusions can only be made under very specific circumstances. In order to claim that this set of predictors explains 66% of the variance in the ratee-related variance (or 55.09% if the exact variance components reported by Jackson et al. [2020] are used), one must know for certain that the predictors do not explain any variance in the other sources of variance of job performance.

This commentary aims to investigate the accuracy of the inferences about the predictive power of the selection tests in predicting ratee-related variance in job performance ratings. For that purpose, a synthetic dataset was created based on the correlation matrix provided by the authors

Table 1. Correlations Between the Selection Tests and Job Performance Ratings

Variable	1	2	3	4	5	6	7	8	9
1. Performance									
2. Interview	.19								
3. EI	.24	.22							
4. GMA	.29	.14	.05						
5. Neuroticism	−.09	−.04	−.47	−.07					
6. Extraversion	.04	.10	.42	.00	−.26				
7. Openness	.05	.04	.33	.19	−.07	.29			
8. Agreeableness	.08	.06	.32	.00	−.16	.20	.19		
9. Conscientiousness	.17	.08	.32	.02	−.23	.19	.16	.29	

Note. EI = emotional intelligence; GMA = general mental ability. From "Selection Tests Work Better Than We Think They Do, and Have for Years" by Foster et al. (2024), *Industrial and Organizational Psychology.*

and the variance components reported by Jackson et al. (2020). Then, the predictors' explained variances in each component were analyzed using this synthetic dataset.

Synthetic data based on the variance components and correlations between the predictors and job performance ratings

In order to test the explained variance in the ratee-relevant variance by the selection tests, synthetic data containing 1,000 cases was created based on the aforementioned statistical information using SPSS Version 29 (IBM Corp, 2023). A total of 12 variables were created: eight predictor variables, one job performance variable, and four variance components for ratee-related variance, rater-related variance, rater–ratee interaction, and variance explained by other factors. The predictor and job performance variables were simulated as standardized variables based on a known correlation matrix with a mean value of zero, a standard deviation of 1, a minimum value of −3, and a maximum value of 3. The simulated job performance variable was used as a stand-in variable for the latent construct of job performance, and the independent variables were simulated by rearranging the regression equation to solve for independent variables using the simulated job performance variable as the dependent variable. The regression coefficients in the rearranged equation were derived from the explained variances in Jackson et al.'s (2020) article.

A linear regression analysis was performed to investigate the predictive power of the eight selection tests on job performance ratings. The combination of eight predictor variables explained 16.6% of the variability in the job performance ratings, similar to the R^2 value of .164 found by the authors of the focal article. The small difference between the two findings is most likely due to the randomness of the generated synthetic dataset. Based on the authors' proposition, compared to the 29.77% ceiling, this corresponds to 55.76% of the variance. Another multiple linear regression analysis was performed using the ratee-related variance component as the dependent variable and the selection tests as the predictors to investigate the accuracy of this finding. The result of this analysis was statistically significant ($F(8, 991) = 13.52, p < .001$), and eight predictors explained 9.8% of the variance in the ratee-relevant variance component. Comparing the simulation-based estimate and the ceiling-based calculation, it can be seen that the latter overestimates the predictive power of the selection tests by a large margin.

Discussion

Considering Foster et al.'s (2024) perspective on the validity of selection tests and the simulation-based evidence, a few key issues need to be addressed to find out how well selection tests work in predicting job performance more accurately. First, a better conceptual understanding of the relationship between selection tests and different variance sources in job performance ratings is warranted. It is likely that conceptually unrelated variance components reduce the observed validity coefficients of the selection tests, but ignoring all sources of variance except for ratee-related variance may lead to an overestimation. The second issue is investigating the relationship between specific selection methods assessing different psychological constructs and various aspects of job performance. Specifically, two key points must be considered regarding this issue: (a) predictive powers of selection tests in predicting broad versus narrow performance dimensions and (b) situational trait relevance based on job-specific work requirements and performance dimensions. Last, interpreting the relationships between correlated predictors and performance indices requires a more nuanced approach, and commonalities between the predictors should be considered meaningful predictors of job-related criteria.

As demonstrated using the synthetic dataset, the percentage of variance explained by the selection tests in ratee-related variance in job performance ratings based on the focal article's authors is too optimistic. Still, the points raised in the focal article point out the need to think more about the relationship between selection test scores and different sources of variance in job performance ratings. Specifically, the arguments made by Foster et al. (2024) challenge the purely empiricist approach in which different sources of variance are treated similarly in terms of their relationships with selection tests. Selection tests aim to measure various knowledge, skills, abilities, and other characteristics of the applicants, and conceptually, they are expected to predict ratee-related variance in job performance ratings. Although the idea that rater-related variance should not be explained by selection tests is valid and definitely sheds light on the nature of the relationship between selection tests and job performance, it is still possible and likely that those predictors explain some variance in rater–ratee interactions and other sources of job performance variance. Overestimating how well the selection test works may make our field conclude that we have reached the fullest of our potential, which is not likely the case. Organizations use selection tests and performance appraisal systems to make critical decisions about selection, promotions, rewards, and terminations, affecting both individual and organizational well-being (O'Neill et al., 2015), and we do not want those decisions to be made with overconfidence. Although selection tests may have been working better than we think they do, they are far from perfect. Therefore, we should still do more research to understand the complex nature of job performance as a psychological construct and clear as much noise from job performance ratings as possible to better understand the relationships between selection tests and job performance.

The current body of knowledge on the nature of job performance suggests that the validity coefficients give us information about how much variability in job performance ratings is explained by selection tests, but it does not necessarily tell us whether the explained variance is shared by all components of the criterion equally. For example, Viswesvaran et al. (2005) discussed the general performance factor, which explains 27% of the variability in the observed ratings and 60% of the variability at the construct level after the corrections. The authors pointed out the possibility that this general factor is explained primarily by general mental ability and conscientiousness. Moreover, Viswesvaran and Ones (2000) also acknowledged the importance of specific factors that apply to particular jobs or across occupations. It can be argued that some selection tests may explain more variance in the generalizable specific factors of job performance, whereas others may explain job-specific factors. For example, a structured interview may explain more variability in Campbell's (1990) oral communication dimension of performance across all occupations. On the other hand, a particular assessment center may explain variability in Borman

and Brush's (1993) crisis-handling aspect of managerial performance. The occupation-specific variability in the job performance ratings due to differential requirements has been addressed by Tett et al. (1999), suggesting that meta-analytical correlations between personality and job performance should be interpreted with caution due to underestimations of the relationships caused by bidirectionality. Specifically, they suggested that some traits, such as agreeableness, may be positively related to job performance for occupations that demand agreeableness (e.g., nurses), whereas it might show negative or no relationships for other jobs with different natures (e.g., law enforcement agencies). These effects may cancel each other out, yielding a lower average mean meta-analytical correlation coefficient, which underestimates the predictive power of the construct. Relatedly, the extent to which agreeableness explains variability in job performance rating will be a function of how beneficial (or detrimental in some cases) the trait for high performance in a given occupation. Further research on job specificity and broad versus narrow performance criteria may help us learn more about the conditions in which the selection tests work better (or worse).

Another vital issue generally overlooked in discussions in the literature about the predictor–criterion relationships is the joint variance between the predictor variables (Schoen et al., 2011). Multiple regression analyses are frequently used to understand how much variability in the outcome variable can be explained when a set of variables is combined, but the results can be misleading when the predictor variables are correlated (Nimon & Oswald, 2013), which is the case for the selection test validity research, especially when constructs and methods are mixed together, as shown in Table 1. Schlaegel et al. (2022) addressed this issue in a study in which they investigated not only the four aspects of emotional intelligence but also the second-, third-, and fourth-order commonalities and found that the joint variances explained unique variances in job performance ratings across three unique samples from Germany, India, and the USA. Similarly, Shuck et al. (2017) demonstrated that when job attitudes were used as predictors of job engagement and its components, common variance shared organizational commitment and job satisfaction, and common variance shared by all three job attitudes explained considerably higher variance in overall engagement ratings as well as the individual subscale scores than the unique variance explained by each attitude. This poses a unique challenge for our field as such commonalities are mostly treated as errors (e.g., common method variance), but it is not necessarily the case all the time as it is also possible that some selection tests have conceptual similarities unrelated to common method variance (Schoen et al., 2011). A good example of this is the Dark Triad personality traits. Multiple studies (e.g., Jones & Figueredo, 2013; Vize et al., 2020) demonstrated that distinct constructs such as Machiavellianism, narcissism, and psychopathy have a shared "core" due to overlapping characteristics such as antagonism. Further research is warranted to explore how to weed out the meaningful commonalities shared by the predictors to better understand their unique predictive contribution on top of the unique effects of the predictors.

Conclusion

The conclusions and recommendations of Foster et al. (2024) provide a new perspective for understanding how good selection tests do in predicting job performance. Although some assumptions they made while recalculating the explained variances in job performance ratings are less likely to occur in realistic scenarios, their idea that rater-related variance should not be affected by the ratee characteristics measured in selection tests suggests that we might be doing better than we think we do but probably not as well as the authors suggested. Given that selection test validity is a critical topic in industrial and organizational psychology for both scholars and practitioners, understanding the true relationships between ratee-related variance in job performance and selection tests is likely to be a treasure hunt we will keep undertaking.

References

Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. *Human Performance*, **6**(1), 1–21. https://doi.org/10.1207/s15327043hup0601_1

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette, & L. M. Hough (Ed.), *Handbook of industrial and organizational psychology* (2nd ed. pp. 687–732). Consulting Psychologists Press.

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

IBM Corp 2023). *IBM SPSS Statistics for Windows*. IBM Corp.

Jackson, D. J. R., Michaelides, G., Dewberry, C., Schwencke, B., & Toms, S. (2020). The implications of unconfounding multisource performance ratings. *Journal of Applied Psychology*, **105**(3), 312–329. https://doi.org/10.1037/apl0000434

Jones, D. N., & Figueredo, A. J. (2013). The core of darkness: Uncovering the heart of the dark triad. *European Journal of Personality*, **27**(6), 521–531.

Nimon, K. F., & Oswald, F. L. (2013). Understanding the results of multiple linear regression: Beyond standardized regression coefficients. *Organizational Research Methods*, **16**(4), 650–674. https://doi.org/10.1177/1094428113493929

O'Neill, T. A., McLarnon, M. J. W., & Carswell, J. J. (2015). Variance components of job performance ratings. *Human Performance*, **28**(1), 66–91. https://doi.org/10.1080/08959285.2014.974756

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, **107**(11), 2040–2068. https://doi.org/10.1037/apl0000994

Schlaegel, C., Engle, R. L., & Lang, G. (2022). The unique and common effects of emotional intelligence dimensions on job satisfaction and facets of job performance: An exploratory study in three countries. *International Journal of Human Resource Management*, **33**(8), 1562–1605. https://doi.org/10.1080/09585192.2020.1811368

Schoen, J. L., DeSimone, J. A., & James, L. R. (2011). Exploring joint variance between independent variables and a criterion: Meaning, effect, and size. *Organizational Research Methods*, **14**(4), 674–695. https://doi.org/10.1177/1094428110381787

Shuck, B., Nimon, K., & Zigarmi, D. (2017). Untangling the predictive nomological validity of employee engagement: Partitioning variance in employee engagement using job attitude measures. *Group & Organization Management*, **42**(1), 79–112. https://doi.org/10.1177/1059601116642364

Tett, R. P., Jackson, D. N., Rothstein, M., & Reddon, J. R. (1999). Meta-analysis of bidirectional relations in personality-job performance research. *Human Performance*, **12**(1), 1–29. https://doi.org/10.1207/s15327043hup1201_1

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, **8**(4), 216–226. https://doi.org/10.1111/1468-2389.00151

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, **90**(1), 108–131. https://doi.org/10.1037/0021-9010.90.1.108

Vize, C. E., Collison, K. L., Miller, J. D., & Lynam, D. R. (2020). The "core" of the dark triad: A test of competing hypotheses. *Personality Disorders: Theory, Research, and Treatment*, **11**(2), 91–99. https://doi.org/10.1037/per0000386