



Commentary on Matching Predictor and Criterion Complexity: The Case of Predicting job Performance by Industrial/Organizational Psychology

Corey E. Miller^{1*}, Suzanne L. Dean², and Michael Brady³

¹Associate Professor, Department of Psychology, Wright State University, 3640 Colonel Glenn Hwy. Dayton, OH 45435, United States.

²Wright State University, 3640 Colonel Glenn Hwy. Dayton, OH 45435, United States.

³Wright State University, 3640 Colonel Glenn Hwy. Dayton, OH 45435, United States.

Article Details

Article Type: Commentary Article

Received date: 03rd April, 2026

Accepted date: 13th April, 2026

Published date: 15th April, 2026

***Corresponding Author:** Corey E. Miller, Ph.D., Associate Professor, Department of Psychology, Wright State University, 3640 Colonel Glenn Hwy. Dayton, OH 45435, United States.

Citation: Miller, C. E., Dean, S. L., & Brady, M., (2026). Commentary on Matching Predictor and Criterion Complexity: The Case of Predicting job Performance by Industrial/Organizational Psychology. *J Ment Health Soc Behav* 8(1):213. <https://doi.org/10.33790/jmhsb1100213>

Copyright: ©2026, This is an open-access article distributed under the terms of the [Creative Commons Attribution License 4.0](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Industrial Organizational Psychology has done a poor job of predicting job performance. We argue that the focus on explanation, construct validity, and broad traits has come at the expense of prediction. We review literature from other subfields of psychology to bolster our argument that I/O Psychology would benefit applying that research to improve prediction of I/O Psychology's main criterion of interest: job performance. The main argument is that in order to predict complex, multi-faceted, heterogeneous criteria, the predictors must also be complex, multi-faceted, and heterogeneous. Although the predictor measures would not achieve traditional psychometric standards of internal consistency, they could still achieve increased prediction.

Key Words: Predictor, Criteria, Job Performance, Validity, construct, Criterion-related, Content, Broad, Narrow, Homogeneous, Heterogeneous

Introduction

Industrial Organizational Psychology has not been very successful at predicting Job Performance

Industrial/Organizational (I/O) psychologists have only done a modest job of predicting performance, even after more than eighty years of research [1]. One of the reasons why we have not been able to better predict performance may be because researchers have emphasized generalizability and explanatory power which leads to singular and homogeneous predictor constructs over predictor constructs that are multiple, narrow, situation-specific, and divergent. Although the simplicity of this research focus may have led to more overarching theories of performance prediction, it likely will not lead to overarching theories that predict performance with any more accuracy than has been achieved to date. The reality is that predicting performance is complex and that better understanding and research can lead to better prediction while maintaining ability to explain the relationships, but there may be a trade-off [2].

Performance is composed of factors are multiple, narrow, and heterogeneous and interact in a way such that the sum is greater than its component parts, or a gestalt. The following research attempted to prove that performance is in fact a gestalt and that a constellation of multiple, narrow, divergent component parts is a better predictor than what has been used traditionally. It would therefore logically follow that , choosing predictor constructs that are a constellation of multiple narrow, situation-specific, divergent constructs is a key way to capture more of the variance in performance and attempt to capture the gestalt. The main focus of this research was focused on determining whether or not there is greater value in using more empirically-derived, heterogeneous predictors than theoretically-derived, homogeneous predictors.

The Consistency Model

Numerous researchers have posited that matching predictors and criteria across multiple dimensions leads to better prediction [3, 4, 5]. Wernimont and Campbell [5] argued that it would be much more beneficial to focus on meaningful samples of behavior, rather than signs of predispositions, as predictors of later performance. Ajzen and Fishbein [3] developed a compatibility principle, concluding that attitude-behavior connections are strongest when an attitude is matched in specificity or generality to behavior. Further, Smith [4] articulated multiple possible sources and measures of performance variation and posited that criteria should parallel predictors in generality and immediacy. More recent research has extended the previous findings to frame of reference effects, which suggest that more work-specific personality measures generally yield stronger relationships with work criteria than general personality [6].

I/O research continues to rely heavily on signs of performance (e.g., selection tests) rather than samples, despite all of the evidence that samples of performance are the most predictive of future performance [5]. Researchers continue to favor the use of broad constructs and broad predictors because of their simplicity and generalizability, which lend themselves easily to theory building.

But simplifying the performance predictor and criterion domain may come at the cost of the ability to predict performance with greater accuracy.

I/O researchers have often focused on measuring one dimension or construct very thoroughly in order to ensure high internal consistency, as opposed to looking at multiple dimensions that differentiate between high and low performers. Researchers have focused on measuring a single dimension in part because singular constructs can be measured more reliably than multiple dimensions in one test, and therefore tend to have higher alpha coefficients. If performance is multidimensional, it stands to reason that a predictor should be multidimensional in order to better represent the construct domain. Advocates of broad constructs argue that a single construct can explicate performance or that performance can be reduced to a single factor [e.g. 7]. We take the contrasting view and suggest that performance is multidimensional and any predictor used to predict performance should also be multidimensional. We briefly review the extensive research on three types of predictors (biodata measures, Situational Judgment Tests, and Assessment Centers) that have been proven conclusively through meta-analytic research to have good predictive validity, but weaker construct validity as they are not measuring a single, well-defined construct.

Situational judgment tests (SJTs) are low-fidelity simulations requiring the respondent to exercise judgment when responding to hypothetical problem situations that occur in work settings. SJTs are typically considered as heterogeneous measurement methods rather than measures of a single construct [7, 8]. The strong validity of SJTs [9] while simultaneously resulting in smaller mean differences among racial subgroups as compared with traditional cognitive ability tests [10, 11, 12] have led to increasing popularity. Similarly, numerous researchers have agreed that Assessment Centers (AC) are also good predictors of job performance [13, 14]. ACs have become widely used because of their predictive validity, and they typically demonstrate less adverse impact than *g* [1]. ACs are thought to be measuring a host of managerial competencies, although they do have some correlation with *g* and personality. Little to no construct validity evidence has been presented in most studies assessing SJT's predictive validity, although they consistently exhibit fairly strong criterion-related validities and smaller racial and sex subgroup differences than other methods, [15]. There has been modest success in improving convergent validity with other measures of similar constructs [15]. SJTs typically have lower internal consistency because they are job-centered as opposed to construct-centered [15, 16]. Assessment centers have demonstrated strong content and criterion-related validity; however, many reviews have suggested that they have weak construct validity [17, 18]. The construct-related validity paradox refers to the assessment center's ability to display relatively satisfactory levels of content and criterion-related validity, but weak construct related validity [13]. A simple explanation of the predictive ability or criterion-related validity is that these measures are focused on sampling what reliably differentiates high and low job performers, similar to work sample tests, that are considered to be the most valid predictors of job performance [14].

Research Supporting Heterogeneous Predictors. The strong criterion-related validity evidence of work samples, background data, situational judgment tests, and assessment centers lead to the conclusion that heterogeneous, job specific predictors such as work samples, background data, situational judgment tests, and assessment centers are useful predictors.

The Overemphasis of Construct Validity

The predominant criticism of heterogeneous tests such as situational judgment tests, assessment centers, and biodata is that they have poor construct validity. Construct validity is the degree to which a test measures what it claims to be measuring [19]. These predictors have

often been described in method-based terms and developed to simulate the job itself as opposed to a specific predictor construct [16], and have evolved independent of the nomological network. These multidimensional predictors are valid predictors of multidimensional jobs which in turn dictates that traditional psychometrics will show low internal consistency reliability and weak construct validity.

The Overemphasis of Internal Consistency Reliability

Some researchers and practitioners have criticized job-specific, heterogeneous predictors because of their low internal consistency reliability. The most commonly used measure of reliability, coefficient alpha, is an inappropriate estimate of the reliability of heterogeneous measures like SJTs [10], work sample tests and background data scales. Measures focusing on background data must include items that cover multiple situations, which in turn limits the magnitude of internal consistency coefficients unless a large number of items (30 or more) are in use [20]. Researchers often decide to report alternate measures of reliability (e.g., test/retest) when dealing with job-specific heterogeneous predictors [e.g., 16].

It is commonly asserted that, all things being equal, higher reliability leads to higher validity and lower reliability leads to lower validity [21]. This assertion is a received doctrine, or a fact that is not to be challenged [22]. Most researchers of heterogeneous predictors attempt to sidestep this issue because classical test theory is rooted so strongly in the idea that reliability underscores validity. It is not logically inconsistent to focus on a complex, multi-dimensional, heterogeneous test domain if one adopts the elegant definition of construct by Binning and Barrett [23], or labels for covarying behaviors. The covarying behaviors encompassed by conscientiousness can be described in terms of behaviors, such as planning ahead, paying attention to detail, and being organized, as one example. Thus, if we define the construct of interest as a behavioral consistency of the behaviors that reliably lead to good job performance, the seemingly logical contradiction is solved. It then follows that internal consistency, or alpha reliability, is not necessary to achieve criterion-related validity, and also follows that construct validity would be a complex, multi-dimensional construct.

Conclusion

Industrial/Organizational Psychology could improve prediction of job performance by applying a different understanding of testing and psychometrics that are common in other sub-fields of psychology. Using Binning and Barrett's [23] definition of construct, better criterion-related validity could be obtained by focusing on behavioral consistency or job-relevant behavior shown to reliably differentiate high and low performers. The explanatory power of these predictors in a logical framework would likely be less clear because they would be heterogeneous, multi-faceted, complex constructs rather than simple, homogeneous, constructs. It also follows that the measures would have lower internal consistency, or alpha reliability than traditionally achieved because they measure more than one dimension. I/O Psychology's overemphasis on internal consistency has led many to believe that criterion-related validity cannot be achieved with a measure of lower internal-consistency; however, the research literature on three of multiple possible examples (biodata, Situational Judgment Tests, and Assessment Centers) is briefly reviewed to rebut this assertion.

Competing Interests: The author declares that they have no competing interests.

References

1. Cascio, W. F., & Aguinis, H. (2008a). Research on industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology*, 93(5), 1062-1081.

2. Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*(1), 40-57.
3. Ajzen, I. & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84*(5), 888-918.
4. Smith, P. (1976). Behaviors, results, and organizational effectiveness: The problem of criteria. In M.D. Dunnette (Ed.) *Handbook of Industrial and Organizational Psychology, 745-775*. Chicago: Rand McNally.
5. Wernimont, P. F. & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*(5), 372-376.
6. Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology, 89*, 150-157.
7. 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.
8. Chan, D. & Schmitt, N. (1997). Video-based versus paper—and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity. *Journal of Applied Psychology, 82*(1), 143-159.
9. McDaniel, M.A. & Nguyen, N.T. (2001). Situational Judgment Tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.
10. McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*(4), 730 – 740.
11. Motowidlo, S. J., Dunnette, M. D., & Carter, G.W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(5), 640-647.
12. Pulakos, E. D., Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9*(3), 241-258.
13. Arthur, Jr., W., Day, E. D., McNelly, T. L., Edens, P. S. (2006). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*(1), 125-153.
14. Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72-98.
15. Ployhart, R. E., & MacKenzie Jr., W. I. (2011). Situational judgment tests: A critical review and agenda for the future. In S. Zedeck (Ed.). *Handbook of Industrial and Organizational Psychology, 2*(2), 237-252.
16. Roth, P. L., Bobko, P., McFarland, L. A., & Buster, M. (2008). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology, 58*, 1009-1037.
17. Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*(6), 1011-1027.
18. Sackett, P. R. & Harris, M. M. (1988). A further examination of the constructs underlying assessment center ratings. *Journal of Business and Psychology, 3*(2), 214-229.
19. Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.
20. Mumford, M. D. & Stokes, G. S. (1992). Developmental determinants of individual action: Theory and practice in applying background measures. In M. D. Dunnette & L. M. Hough (Eds.). *Handbook of Industrial and Organizational Psychology, 3*(2), 61-138.
21. Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. New York: McGraw-Hill.
22. Barrett, G. V. (1972). Research models of the future for industrial and Organizational psychology. *Personnel Psychology, 25*, 1-18.
23. Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis and evidential bases. *Journal of Applied Psychology, 74*, 478-494