


COMMENTARY

Revisiting predictor–criterion construct congruence: Implications for designing personnel selection systems

Leaetta M. Hough¹  and Frederick L. Oswald²

¹The Dunnette Group Ltd., St. Paul, MN, USA and ²Rice University, Department of Psychological Sciences, Houston, TX, USA

Corresponding author: Leaetta Hough; Email: leaetta@msn.com

Keywords: Personnel selection; Synthetic validity; Construct congruence

Overview

In their focal article, Sackett et al. ([in press](#)) describe implications of their new meta-analytic estimates of validity of widely used predictors for selection of employees. Contradicting the received wisdom of Schmidt and Hunter (1998), Sackett et al. conclude that predictor methods with content specifically tailored to jobs generally have greater validity for predicting job performance than general measures reflecting psychological constructs (e.g., cognitive abilities, personality traits). They also point out that standard deviations around the mean of their meta-analytic validity estimates are often large, leading to their question “why the variability?” (p. x). They suggest many legitimate contributors.

We propose an additional moderator variable of critical importance: *predictor-criterion construct congruence*, accounting for a great deal of variability in validity coefficients found in meta-analysis. That is, the extent to which *what is measured* is congruent with *what is predicted* is an important determinant of the level of validity obtained. Sackett et al. (2022) acknowledge that the strongest predictors in their re-analysis are job-specific measures and that a “closer behavioral match between predictor and criterion” (p. 2062) might contribute to higher validities. Many in our field have also noted the importance of “behavioral consistency” between predictors and criteria relevant to selection, while also arguing for another type of congruence: the relationships between constructs in both the predictor and criterion space (e.g., Bartram, 2005; Campbell et al., 1993; Campbell & Knapp, 2001; Hogan & Holland, 2003; Hough, 1992; Hough & Oswald, 2005; Pulakos et al., 1988; Sackett & Lievens, 2008; Schmitt & Ostroff, 1986).

The above reflects an important distinction between two types of congruence: *behavior-based* congruence and *construct-based* congruence. When ‘past behavior predicts future behavior’ (as might be possible for jobs requiring past experience and where behavior-oriented employment assessments such as interviews, biodata, and work samples are involved), behavior-based congruence exists. Behavior-based assessments can vary a great deal across jobs but tend to ask about past experiences that are influenced by a complex mix of KSAOs. By contrast, construct-based congruence aligns employment tests of job-relevant KSAOs (e.g., verbal and math skills, conscientiousness) with relevant work criteria, such as technical performance or counterproductive work behavior (e.g., Campbell & Wiernik, 2015).

What we are suggesting strongly here is that regardless of the approach to congruence adopted in selection, it is the congruence between predictor and criterion constructs that is a key factor

Revised and submitted April 10, 2023, to *Industrial and Organizational Psychology: Perspectives in Science and Practice* as a commentary to focal article entitled “Revisiting the Design of Selection Systems in Light of New Findings Regarding the Validity of Widely Used Predictors” by Sackett, Zhang, Berry, and Lievens.

influencing the levels of validity found across all personnel selection assessment tools in the Sackett et al. (2022) meta-analysis. I–O psychologists have internalized that methods are not the same as constructs (Arthur & Villado, 2008); virtually *any* KSAO can be measured through methods such as the interview, biodata, and work samples. If the conclusion is “the structured interview tends to be the most valid predictor in selection contexts,” it is not *only* because of the structure. Instead, it is ideally due to behavior-based congruence, such that what is being measured in the structured interview is tapping the same KSAOs that are measured in the performance criterion. This congruence is much more likely to happen when a selection system is developed based on job analysis of relevant KSAOs (Morgeson et al., 2019; Steel et al., 2006); it is much less likely to happen otherwise. A lack of predictor–criterion congruence is an important source of error contributing to meta-analytic correlations, where correlations could increase (due to method variance such as likability in both the interview and in performance ratings; Schmitt et al., 1996) or correlations could decrease (due to KSAO incongruence).

In an ever-changing world of work, understanding relationships between predictor and criterion constructs is key to developing highly valid selection systems quickly; validities are useful when they can be transportable and generalizable to new job. In addition to what can be properly inferred from meta-analyses such as Sackett et al. (2022), synthetic validation is the foundation upon which such systems can be built. Highly valid selection systems can be designed intentionally, efficiently, and effectively by reviewing meta-analytic validity evidence to synthetically estimate validity for the situation at hand. In a world where selection systems are subject to government regulations, and social justice requires that those systems be fair to all, strategies that rely on known inputs and produce known outcomes such as through predictor–criterion construct congruence as enabled through synthetic validity strategies are key to successful design of selection systems. Artificial intelligence talent management solutions run the risk of black-box empiricism, where the data and/or the underlying algorithm might be opaque and resistant to the construct-level understanding afforded by congruence-based approaches.

In the pages that follow, we provide representative evidence supporting our contention that predictor–criterion congruence is an important determinant of the magnitudes of validity.

Interview: structured and unstructured

Sackett et al. found that the operational mean validity of the unstructured interview is 0.19, whereas the validity of structured interviews is 0.42, the highest level of validity of all the predictors they studied. Huffcutt et al. (2001) examined meta-analytically the constructs measured in employment interviews and stated that “at least part of the reason why structured interviews tend to have higher validity is because they focus more on constructs that have a stronger relationship with job performance.” (p. 897). Our point precisely: When there is a correspondence between the substance of what is measured by the criterion measure and the substance of what is measured by the predictor measure, validity will be higher. For example, the critical incident approach to job analysis results in constructs based on work behaviors that readily become the basis for developing behavioral-based and situation–judgment interviews. Both types of interviews are structured interviews, and both have high criterion-related validity (e.g., Huffcutt et al., 2014).

Empirically keyed biodata

Sackett et al. (in press) report the mean validity of empirically keyed biodata inventories as 0.38 and the mean validity of rationally keyed biodata as 0.22. These updated findings are based on the meta-analysis of Speer et al. (2022), who explained that “when biodata scores were correlated with theoretically aligned performance ratings, rational scoring resulted in similar validity coefficients as empirical scoring” (p. 1678). Hough and Paullin (1994) also arrived at a similar conclusion

when comparing biodata measures developed by rational (theoretical) and empirical means against the same criteria. In short, when predictor–criterion construct congruence is high, validity coefficients are not only higher, they stand to be more consistent across samples (validity coefficients that are less likely to capitalize on chance).

Work samples and assessment centers

Sackett et al. (in press) report the mean validity of work samples as 0.33. The updated, appropriately corrected mean validity of assessment center was also reported as 0.33. Clearly, the development of work samples is based on an understanding of the job and the requirements of the work, and Sackett et al., point out that assessment centers are work samples for managerial jobs. It goes without saying, but we will: predictor–criterion construct congruence is high for both work samples and assessment centers.

Situational judgment tests

Both types of situational judgment tests for which Sackett et al. (in press) separately report validities, i.e., knowledge and behavioral tendency (validities of 0.26 for both), are also typically developed to measure job requirements and work situations that are based on job/work analyses. As noted above, situational judgment tests and behavior-based (structured) interviews are often developed using the critical incident method for job analysis and can be considered job simulations, albeit low-fidelity job simulations (McDaniel & Nguyen, 2001; Motowidlo et al., 1990). Like us, others also suggest that one reason situational judgments sometimes do not correlate with performance is lack of predictor–criterion congruence (e.g., Whetzel & Reeder, 2016). One example of evidence for this point of view comes from research on the criterion-related validity of situational judgment tests for predicting performance in medical school. Such research indicates that interpersonally oriented situation judgment tests predict performance in patient care-oriented medical schools, but not in basic science-oriented medical schools (Lievens et al., 2005). This finding allows us to echo our general point: The greater the predictor-criterion construct congruence, the higher the validity.

General mental ability (GMA)

Perhaps the most important finding in the Sackett et al., meta-analysis is that the mean validity coefficient for general mental ability (GMA) tests predicting overall job performance is lower than that from previous meta-analyses. Sackett et al. suggest one possibility for these lower validities is due to the greater importance of interpersonal skills and team-based work of today's work, as compared with manufacturing type jobs included in older meta-analyses. If that is the case, this of course supports our thematic point that the requirements of the work need to be understood and predictor–criterion constructs aligned. Nye et al. (2022) meta-analyzed the differential validity of narrow cognitive abilities for predicting diverse criteria. Although they found incremental validity of narrow cognitive abilities over general mental ability, they found that the match between the narrow ability and job tasks did not have a *substantial* effect on validity. They state that a possible explanation is that the “*breadth* of each of the specific job performance dimensions assessed (e.g., task performance or organizational citizenship behavior) was incompatible with the narrower cognitive abilities” (p. 1136). Perhaps the complex performance of employees—and the overly simplistic ratings of their supervisors—is often unavoidable in an organization. We hope this state of affairs does not discourage efforts at measurement of criteria better reflecting specific abilities, shedding greater light on the benefits of measuring specific versus general abilities.

Personality variables

Nowhere in the research literature is there more evidence of the importance of predictor–criterion construct congruence than in the study of personality variables. First, Sackett et al. (*in press*) separate personality measures of variables, such as Conscientiousness, according to those measures that are contextualized for work settings and those that are not. In a related vein, evidence that validities are higher for personality variables that are theoretically linked to the criteria has been shown for managerial work (Bartram, 2005) as well as for a variety of other jobs and criterion constructs including teamwork, counterproductive behavior, organizational citizenship, creativity, and work engagement (e.g., Hogan & Holland, 2003; Hough et al., 1990; Hough & Oswald, 2021; Oswald & Hough, 2008). Research on predictor–criterion construct congruity at the facet level of personality (more refined than the Big Five) continues due to producing informative patterns of convergent and discriminant validity (Hough & Oswald, 2005; Hough & Johnson, 2013; Judge et al., 2013; Steel et al., 2019). To advance this work we need to continue investigating refined taxonomic structures of personality, as the literature is now producing and beginning to align with work-related outcomes (e.g., Soto et al., 2022). As Schneider et al. (1996) stated eloquently, “Increasing use of narrower personality and job performance constructs, in concert with construct-oriented methodology, will greatly enrich I-O personality research. If we limit ourselves only to broad traits and general laws, we may find that we have fatally cut ourselves on the blade of Occam’s razor” (p. 653).

Implications for personnel selection and I-O psychology

Sackett et al. have provided a very important correction to our understanding of the level of validity of widely used predictors in personnel selection systems. However, rank ordering predictors according to their level of validity for predicting overall job performance is somewhat unhelpful, especially if very important moderators of predictor–criterion validities are left on the table (e.g., jobs, settings, samples). Sackett et al. also suggest caution when comparing findings across meta-analyses without clear understanding of the specific components underlying performance ratings—as we outlined in the behavior-based and construct-based congruence approaches to selection.

Although we can never carve nature at its exact joints, we can build useful hierarchical taxonomies of constructs at different levels of refinement, such as that found in the metaBUS (www.metabus.org; Bosco et al., 2015; Bosco et al., 2017) extensive organizational database of correlational relationships between predictors and criteria. Complementing this taxonomic structure could be the nomological-web clustering of predictor-criterion relationships, as described by Hough et al. (2015) for personality variables. Taxonomies allow us to use meta-analysis and synthetic validity in flexible ways to estimate the utility of newly developed prediction systems for new and newly configured jobs. Although selection researchers have yet to adopt a standard taxonomy such as that offered by metaBUS, we still are further along in understanding predictor and criterion constructs, relationships, and pertinent moderators at a more refined level than ever before. And we are ready to do that, because as I-O psychologists, we now embrace the criterion as multidimensional, so much so that it is almost hard to appreciate that when Guion (1961) and Dunnette (1963) admonished us against focusing on the “ultimate” criterion, there was a reason. We still need to continue pushing ahead in refining our thinking, our constructs, and our selection research even further. We should not rest our laurels on meta-analytic estimates that come with an unavoidable heterogeneity of samples and settings—and last but not least, jobs.

References

- Arthur, W., Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, *93*, 435–442.
- Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology*, *90*, 1185–1203.
- Bosco, F. A., Steel, P., Oswald, F. L., Uggerslev, K. L., & Field, J. G. (2015). Cloud-based meta-analysis to bridge science and practice: Welcome to metaBUS. *Personnel Assessment and Decisions*, *1*, 3–17.
- Bosco, F., Uggerslev, K., & Steel, P. (2017). MetaBUS as a vehicle for facilitating meta-analysis. *Human Resource Management Review*, *27*, 237–254.
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman, (Eds.), *Personnel Selection in Organizations* (pp. 35–70). Jossey Bass.
- Campbell, J. P., & Knapp, D. J. (Eds.) (2001). *Exploring the limits in personnel selection and classification*. Erlbaum
- Campbell, J. P., & Wiernik, B. M. (2015). The modeling and assessment of work performance. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 47–74.
- Dunnette, M. D. (1963). A note on THE criterion. *Journal of Applied Psychology*, *47*, 251–254.
- Guion, R. M. (1961). Criterion measurement and personnel judgments. *Personnel Psychology*, *14*, 141–149.
- Hogan, J., & Holland, B. (2003). Using theory to evaluate personality and job-performance relations: A socioanalytic perspective. *Journal of Applied Psychology*, *88*, 100–112.
- Hough, L. M. (1992). The “Big Five” personality variables – construct confusion: Description versus prediction. *Human Performance*, *5*, 139–155.
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities [Monograph]. *Journal of Applied Psychology*, *75*, 581–595.
- Hough, L. M., & Johnson, J. W. (2013). Use and importance of personality variables in work settings. In I. B. Weiner (Ed.-in-Chief) N. Schmitt & S. Highhouse (Vol. Eds.), *Handbook of psychology* (pp. 211–243). Wiley.
- Hough, L. M., & Oswald, F. (2005). They’re right, well . . . mostly right: Research evidence and an agenda to rescue personality testing from 1960s insights. *Human Performance*, *18*, 373–387.
- Hough, L. M., Oswald, F. L., & Ock, J. (2015). Beyond the Big Five: New directions for personality research and practice in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*, 183–209.
- Hough, L. M., & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research and use of biographical information in selection and performance Prediction* (pp. 109–145). Consulting Psychologists Press.
- Hough, L. M., & Oswald, F. L. (2021). Personality determinants of engagement. In J. P. Meyer & B. Schneider (Eds.), *Research Agenda for Employee Engagement in a Changing World of Work* (pp. 67–85). Elgar Publishing.
- Huffcutt, A. I., Conway, J. M., Roth, P. L., & Stone, N. J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, *86*, 897–913.
- Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2014). Moving forward indirectly: Reanalyzing the validity of employment interviews with indirect range restriction methodology: Employment interview validity. *International Journal of Selection and Assessment*, *22*, 297–309.
- Judge, T. A., Rodell, J. B., Klinger R., & Simon, L. S. (2013). Hierarchical representations of the five-Factor Model of Personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*, 875–925.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, *90*, 442–452.
- 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.
- Morgeson, F. P., Brannick, M. T., & Levine, E. L. (2019). *Job and work analysis: Methods, research, and applications for human resource management*. Sage Publications.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, *75*, 640–647.
- Nye, C. D., Ma, J., & Wee, S. (2022). Cognitive ability and job performance: Meta-analytic evidence for the validity of narrow cognitive abilities. *Journal of Business and Psychology*, *37*, 1119–1139.
- Oswald, F. L., & Hough, L. M. (2008). Personality testing and industrial-organizational psychology: A productive exchange and some future directions. *Industrial and Organizational Psychology*, *1*, 323–332.
- Pulakos, E. D., Borman, W. C., & Hough, L. M. (1988). Test validation for scientific understanding: Two demonstrations of an approach to studying predictor-criterion linkages. *Personnel Psychology*, *41*, 703–716.
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology*, *59*, 419–450.

- 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**, 2040–2068.
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F.** (in press). *Revisiting the design of selection systems in light of new findings regarding the validity of widely used predictors* **16**(3), 283–300. Industrial and Organizational Psychology.
- Schmidt, F. L., & Hunter, J. E.** (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, **124**(2), 262–274.
- Schmitt, N., Pulakos, E. D., Nason, E., & Whitney, D. J.** (1996). Likability and similarity as potential sources of predictor-related criterion bias in validation research. *Organizational Behavior and Human Decision Processes*, **68**(3), 272–286.
- Schmitt, N., & Ostroff, C.** (1986). Operationalizing the ‘behavioral consistency’ approach: Selection test development based on a content-oriented strategy. *Personnel Psychology*, **39**, 91–108.
- Schneider, R. J., Hough, L. M., & Dunnette, M. D.** (1996). Broad-sided by broad traits: How to sink science in five dimensions or less. *Journal of Organizational Behavior*, **17**, 639–655.
- Soto, C. J., Napolitano, C. M., Sewell, M. N., Yoon, H. J., & Roberts, B. W.** (2022). An integrative framework for conceptualizing and assessing social, emotional, and behavioral skills: The BESSI. *Journal of Personality and Social Psychology*, **123**(1), 192–222.
- Speer, A. B., Tenbrink, A. P., Wegmeyer, L. J., Sendra, C. C., Shihadeh, M., & Kaur, S.** (2022). Meta-analysis of biodata in employment settings: Providing clarity to criterion and construct-related validity estimates. *Journal of Applied Psychology*, **107**, 1678–1705.
- Steel, P. D. G., Huffcutt, A. I., & Kammeyer-Mueller, J.** (2006). From the work one knows the worker: A systematic review of the challenges, solutions, and steps to creating synthetic validity. *International Journal of Selection and Assessment*, **14**, 16–36.
- Steel, P., Schmidt, J., Bosco, F., & Uggerslev, K.** (2019). The effects of personality on job satisfaction and life satisfaction: A meta-analytic investigation accounting for bandwidth-fidelity and commensurability. *Human Relations*, **72**, 217–247.
- Whetzel, D. L., & Reeder, M. C.** (2016). Why some situational judgment tests fail to predict job performance (and others succeed). *Industrial and Organizational Psychology*, **9**, 71–77.