Industrial and Organizational Psychology (2026), **19**, 219–223
doi:10.1017/iop.2026.10081

CAMBRIDGE UNIVERSITY PRESS

COMMENTARY

The devil is in the details: Methodological nuances and challenges in evaluating construct redundancy

In-Sue Oh[1] and Huy Le[2]

[1]Department of Management, Fox School of Business, Temple University, USA and [2]Department of Management, Alvarez College of Business, University of Texas at San Antonio, USA
Corresponding author: In-Sue Oh; Email: insue.oh@gmail.com

The focal article by Bowling et al. (2026) offers a timely and necessary critique of construct proliferation in industrial and organizational (I-O) psychology. The authors call for a temporary moratorium on introducing new constructs and provide a checklist (Table 1) intended to ensure adequate scrutiny of potential construct proliferation, which includes multiple steps for examining both theoretical and empirical distinctiveness. Although the checklist is helpful and largely well grounded, the proposed steps for gauging empirical distinctiveness (or empirical redundancy) warrant additional discussion and caution to be more informative and practically useful. In this commentary, we highlight methodological nuances and challenges involved in assessing empirical redundancy, particularly focusing on (a) the suggested construct-level correlation cutoff of .85 as a universal, uncontextualized benchmark for potential construct redundancy, (b) the information required to estimate construct-level correlations, (c) the importance of broader nomological network evidence, and (d) the role of sampling error.

Unknown bases for the construct-level r of .85 as a sign of empirical redundancy

Step 1 under "empirical distinctiveness" in the construct proliferation checklist (Table 1 of the focal article) asks whether a study reports correlations between a new *construct* and theoretically similar *constructs*. Regarding this step, Bowling et al., suggest that "construct-level correlations of .85 or higher may indicate significant construct redundancy." The focus on construct-level correlations in this suggestion is consistent with the focal article's broader point that construct redundancy often persists because most studies report only observed *measure*-level correlations that can be substantially attenuated unless corrected for all major sources of measurement error, giving a false impression that the underlying constructs may not be highly correlated and thus likely to be distinct (Banks et al., 2018; Le et al., 2009; Schmidt et al., 2003).[1]

However, a universal, uncontextualized construct-level cutoff, without explicit guidance about how it should be adjusted depending on the construct in question and its reliability (the generalized coefficient of equivalence and stability [GCES]; see our discussion in the next section) risks being interpreted too mechanically and doing more harm than service to the field of I-O

Enquiries concerning this article should be sent to In-Sue Oh, Department of Management, Fox School of Business, Temple University, 1801 Liacouras Walk, Philadelphia, PA 19122. Email: insue.oh@temple.edu. Tel: (215) 204-4112. This study was supported by Temple University's sabbatical award granted to In-Sue Oh.

[1]Another important source of attenuation that can create the illusion of construct distinctiveness is range restriction (see Schmidt & Hunter, 2015, Ch. 3 for details).

© The Author(s), 2026. Published by Cambridge University Press on behalf of Society for Industrial and Organizational Psychology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/iop.2026.10081 Published online by Cambridge University Press

psychology (see Bosco et al., 2015 for similar concerns with universal effect size benchmarks). As far as we are concerned, the .85 guideline is not necessarily wrong as a heuristic, but its usefulness at least depends on (a) the availability and quality of the needed reliability (in most cases, GCES) information and (b) the construct content domain, which can influence the magnitude of measurement error and resulting reliability (Le et al., 2009; Schmidt et al., 2003). For the recommended cutoff to be meaningful and consequently widely accepted, the authors of the focal article should have provided empirical and/or theoretical evidence justifying why they chose the value of .85. Although we believe that this cutoff is only a rule of thumb, there should be sufficient caution about its use. As discussed in the following sections, there are many factors that potentially influence estimated construct-level correlations, which may limit the usefulness of a single universal, uncontextualized cutoff value signifying the empirical redundancy between psychological and organizational constructs.

Challenges in obtaining information needed to estimate the construct-level correlations

Step 3 under "empirical distinctiveness" in the checklist clarifies that studies need to be designed to account for all four major sources of measurement error (random, item-specific, scale-specific, and transient errors). That is, the most relevant form of reliability is the GCES (Le et al., 2009), which allows the construct-level r to be estimated as:

$$\text{observed } r \,/\, \sqrt{GCES_{new\ construct} \times GCES_{existing\ construct}}.^{2}$$

However, the challenge is that GCES is rarely available in the literature, given the complexity in the study design that allows for its estimation as discussed in the following (see Le et al., 2009 for more details). A rigorous design that allows for the estimation of GCES requires at least two measurement occasions and at least two parallel measures for each construct to separate key measurement error components/sources (Le et al., 2009). This design is costly and time consuming, and thus rarely implemented (see Le et al., 2010; Le & Pan, 2021 for very rare exceptions), so GCES is rarely reported and mostly unknown for most psychological and organizational constructs—precisely the practical obstacle in implementing Steps 1 and 2 in the checklist (i.e., estimating construct-level correlations and the incremental validity of the new construct in question over other similar constructs).

For example, the construct-level incremental validity analysis (Step 2) requires a construct-level correlation matrix including all the variables involved, all fully corrected for measurement error using GCES (Steps 1 and 3). That is, researchers need to conduct a very complex study in which all study constructs including a newly proposed and theoretically similar constructs are measured across at least two measurement occasions and with at least two parallel measures to obtain GCES values for all study variables (Step 3), their construct-level correlation matrix (Step 1) as input for construct-level incremental validity (Step 2), and other relevant construct-level analyses (Step 4; see Figures 1 and 2 in Le & Pan, 2021). To be clear, our point here is not to discourage the implementation of these steps (Steps 1–4) in the checklist but to raise awareness of the complexity and difficulty associated with them.

One partial workaround is to use reliability generalization evidence to inform what GCES values tend to look like within a construct content domain. For example, in the personality domain, Gnambs (2015) synthesized various types of reliability evidence for each Big Five trait via

[2]Estimating the construct-level correlation using this disattenuation formula is the same as using structural equation modeling, as demonstrated in Le et al. (2009; also see Le et al., 2010 and Le & Pan, 2021 for how to specify such models).

https://doi.org/10.1017/iop.2026.10081 Published online by Cambridge University Press

meta-analysis and derived GCES values indirectly[3] by updating Pace and Brannick's (2010) initial meta-analytic work. The meta-analytic GCES values for the Big Five traits vary between .49 (Agreeableness) and .67 (Neuroticism). This indicates that GCES values can be considerably smaller than typically assumed and vary quite a bit depending on the construct in question.

This has a direct implication for the .85 cutoff because the observed correlation required to yield a construct-level correlation of .85 depends considerably on relevant reliability (GCES) values. The unavailability of GCES makes the .85 cutoff hard to interpret and trust because the same construct-level correlation (e.g., .85) implies different observed correlations depending on the assumed GCES values, which are also likely to vary across constructs like any other form of reliability (e.g., coefficient alpha) and as demonstrated in Gnamb's (2015) reliability generalization study. That is, without knowing (or estimating) the relevant reliability (GCES) and whether it is consistent across constructs, it is impossible to know what level of observed correlation triggers concern about potential construct redundancy. Likewise, it is difficult, if not impossible, to justify the single, across-the-board construct-level cutoff of .85 suggested in the focal article.

A missing empirical step: Construct-level nomological network evidence

The most decisive evidence for empirical distinctiveness vs. redundancy is not a single construct-level correlation in isolation, but the extent to which the new construct and other theoretically similar constructs show meaningfully different patterns of relations within their broader nomological networks (Cronbach & Meehl, 1955; Le & Pan, 2021; Shaffer et al., 2016). Importantly, Bowling et al.'s checklist in Table 1 includes a nomological network criterion under "theoretical and conceptual distinctiveness" (i.e., whether the new construct's hypothesized pattern of relations with external variables is distinct). Our concern is that although the "empirical distinctiveness" portion of the checklist foregrounds (a) construct-level correlations with similar constructs, (b) construct-level incremental validity, and (c) study designs that account for all major sources of measurement error, it fails to include a nomological network test as a clearly articulated *empirical* step parallel to other steps. A construct-level incremental validity test (Step 2)—in which X_1 (new construct) adds prediction over X_2 (existing construct) for outcome Y—is informative, but it is narrower than asking whether X_1 and X_2 exhibit meaningfully different patterns of relations across a set of antecedents, correlates, moderators/mediators, and criteria at the construct level.

A simple demonstration underscores why correlation alone is insufficient in gauging empirical redundancy. For example, consider a hypothetical scenario where two constructs (a new construct, X_1 and an existing construct, X_2) are extremely highly correlated (e.g., $r_{12} = .95$) and the existing construct relates to a third variable, X_3 at $r_{23} = .30$. Under that situation, the correlation between the new construct and the third variable, r_{13} can still depart from .30 widely, ranging from −.01 to .58 (McCornack, 1956). This example clearly illustrates that even a very strong construct-level correlation (e.g., .95) does not, by itself, determine whether the constructs are empirically redundant in the broader network.[4]

[3]Based on Le et al. (2009), Gnambs (2015) derived GCES indirectly by combining meta-analytic estimates of the generalized coefficient of equivalence (GCE) and transient error variance (TEV), such that GCES was estimated as GCE−TEV. GCE represents the correlation between scores from different measures/scales intended to assess the same construct when administered on the same occasion, thereby accounting for random, item-specific, and scale-specific errors. TEV reflects occasion-specific (transient) fluctuation in observed scores, such as short-term mood or situational influences, that can differ across measurement occasions. In this approach, TEV can be approximated as the difference between the coefficient of equivalence (CE) and the coefficient of equivalence and stability (CES) (see Le et al., 2009 and Schmidt et al., 2003 for more details). Gnambs' meta-analysis shows that the GCE for the Big Five traits ranged from .62 (Agreeableness) to .76 (Neuroticism), and TEV ranged between 8% (Extraversion) and 13% (Agreeableness) of the observed test score variances.

[4]Note that if r_{12} is .85 (the suggested cutoff value), r_{13} can vary even more widely, from −.25 to .76.

https://doi.org/10.1017/iop.2026.10081 Published online by Cambridge University Press

Consistent with this logic, only a few studies that have explicitly attempted to correct for all major sources of measurement error while evaluating empirical redundancy have indeed relied on converging nomological evidence. For instance, Le et al. (2010) reported (a) a very high construct-level relation (.91) between job satisfaction and organizational commitment and (b) their broadly similar construct-level relations with affective traits. Across two samples, Le and Pan (2021) similarly reported very high construct-level relations (.83 to .94) among perceived organizational justice dimensions (i.e., overall, distributive, procedural, and interpersonal justice) alongside their broadly similar construct-level relations with multiple theoretically relevant antecedents, correlates, and outcomes.[5] These example studies, although their study design is complex, reinforce that empirical redundancy claims are most strongly supported when (a) construct-level correlations are strong and (b) the patterns of nomological networks converge at the construct level.

Another caveat: Sampling error and the fragility of the results from small samples

The estimation of construct-level correlations and all related analyses mentioned above, such as incremental validity and nomological network comparisons, depend critically on adequate sample size and representativeness. Sampling error will be amplified in corrected correlations (construct-level correlations), and that distorting effect is especially pronounced when sample sizes are small. Viswesvaran et al. (2014) emphasized that the accuracy of psychometric corrections is maximized when sampling error is diminished primarily by aggregating across many samples via psychometric meta-analysis rather than relying on single-sample corrections. Accordingly, a single small-sample study's construct-level correlation meeting or exceeding .85 and other subsequent analyses should not be treated as sufficient evidence of empirical redundancy. Along this line, a practical addition to the empirical checklist would be an explicit step asking whether conclusions about empirical redundancy are supported by (a) large-sample evidence, (b) replications across independent samples, and/or (c) meta-analytic evidence (see Bosco et al., 2015 for an example).

Summary and conclusion

In summary, we commend Bowling et al., for providing an essential framework and a potentially useful checklist for addressing construct proliferation. At the same time, the checklist's empirical redundancy guidance, in our view, appears to oversimplify the nuances and challenges in evaluating empirical redundancy, an important sign or indicator of construct proliferation. In particular, the recommendation to use a single, uncontextualized construct-level correlation cutoff of .85 is difficult to interpret and trust without explicit guidance on and attention to how strongly the cutoff depends on assumed GCES values. Moreover, although nomological network distinctiveness appears in the checklist as a step for checking conceptual redundancy (vs. distinctiveness), it would be beneficial to emphasize construct-level nomological network tests, particularly given its criticality, as an explicit step for checking empirical redundancy alongside other relevant steps. Finally, because construct-level (corrected) correlations can be fragile in small samples, sampling error and the value of replication/meta-analytic aggregation deserve explicit attention in the checklist.

In conclusion, we agree with Bowling et al., that construct proliferation is a long overdue issue requiring greater attention in I-O psychology. The focal article is no doubt a major and much needed step forward. However, we would like to raise awareness that the methodological steps needed to assess empirical redundancy are often more nuanced and complex than the focal article and the enclosed checklist may suggest—the devil is in the details!

[5]As shown in Le et al. (2021), researchers can examine whether "their patterns of correlations with other constructs (are consistent or different) using hierarchically nested CFA models where equality constraints are imposed on relevant correlations (cf. Le et al., 2010)."

https://doi.org/10.1017/iop.2026.10081 Published online by Cambridge University Press

References

Banks, G. C., Gooty, J., Ross, R. L., Williams, C. E., & Harrington, N. T. (2018). Construct redundancy in leader behaviors: A review and agenda for the future. *Leadership Quarterly*, **29**(1), 236–251. https://doi.org/10.1016/j.leaqua.2017.12.005

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, **100**(2), 431–449. https://doi.org/10.1037/a0038047

Bowling, N. A., Sessa, V. I., Shaffer, J. A., & Banks, G. C. (2026). Read my lips: No new constructs! Construct proliferation as a threat to the future of I-O psychology. *Industrial & Organizational Psychology*, **19**(2), 186–202.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, **52**(4), 281–302. https://doi.org/10.1037/h0040957

Gnambs, T. (2015). Facets of measurement error for scores of the Big Five: Three reliability generalizations. *Personality & Individual Differences*, **84**, 84–89. https://doi.org/10.1016/j.paid.2014.08.019

Le, H., & Pan, L. (2021). Examining the empirical redundancy of organizational justice constructs. *Organizational Behavior & Human Decision Processes*, **163**, 195–217. https://doi.org/10.1016/j.obhdp.2021.03.004

Le, H., Schmidt, F. L., Harter, J. K., & Lauver, K. J. (2010). The problem of empirical redundancy of constructs in organizational research: An empirical investigation. *Organizational Behavior & Human Decision Processes*, **112**(2), 112–125. https://doi.org/10.1016/j.obhdp.2010.02.003

Le, H., Schmidt, F. L., & Putka, D. J. (2009). The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organizational Research Methods*, **12**(1), 165–200. https://doi.org/10.1177/109442810730

McCornack, R. L. (1956). A criticism of studies comparing item-weighting methods. *Journal of Applied Psychology*, **40**(5), 343–344. https://doi.org/10.1037/h0045635

Pace, V. L., & Brannick, M. T. (2010). How similar are personality scales of the "same" construct? A meta-analytic investigation. *Personality & Individual Differences*, **49**(7), 669–676. https://doi.org/10.1016/j.paid.2010.06.014

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, **8**(2), 206–224. https://doi.org/10.1037/1082-989X.8.2.206

Shaffer, J. A., DeGeest, D., & Li, A. (2016). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods*, **19**(1), 80–110. https://doi.org/10.1177/1094428115598239

Viswesvaran, C., Ones, D. S., Schmidt, F. L., Le, H., & Oh, I.-S. (2014). Measurement error obfuscates scientific knowledge: Path to cumulative knowledge requires corrections for unreliability and psychometric meta-analyses. *Industrial & Organizational Psychology*, **7**(4), 507–518. https://doi.org/10.1111/iops.12186

Cite this article: Oh, I-S. & Le, H. (2026). The devil is in the details: Methodological nuances and challenges in evaluating construct redundancy. *Industrial and Organizational Psychology* **19**, 219–223. https://doi.org/10.1017/iop.2026.10081

https://doi.org/10.1017/iop.2026.10081 Published online by Cambridge University Press