


COMMENTARY

Is it also time to revisit situational specificity?

Justin Angermeier DeSimone and Tyler Nicole Abayon Fezzey 

The University of Alabama, Tuscaloosa, AL, USA

Corresponding author: Justin Angermeier DeSimone; Email: JADW27@gmail.com

Sackett et al.'s (2023) focal article asserts that the predictors with the highest criterion-related validity in selection settings are specific to individual jobs and emphasizes the importance of adjusting for range restriction (and attenuation) using study-specific artifact estimates. These positions, along with other recent perspectives on meta-analysis, lead us to reassess the extent to which situational specificity (SS) is worth consideration in organizational selection contexts. In this commentary, we will (a) examine the historical context of both the SS and validity generalization (VG) perspectives, (b) evaluate evidence pertaining to these perspectives, and (c) consider whether it is possible for both perspectives to coexist.

Situational specificity and validity generalization

Until the mid-1970s, SS was a predominant paradigm in organizational scholarship (Guion, 1965; Lawshe, 1948). For example, Albright et al. (1963, p. 18) argued that “jobs that seem the same from one place to another often differ in subtle yet important ways.” Ghiselli (1966, p. 28) claimed that “a given test applied to workers on a given job is very likely to have greater validity in one organization than in another.” Ghiselli recognized the influence of sampling error and testing conditions on this variation in validity but argued that “differences in the nature and the requirements for nominally the same job in different organizations” accounted for a large portion of variation (p. 28).

Nearly a half century ago, Frank Schmidt, Jack Hunter, and their colleagues challenged SS by introducing an alternative (and allegedly contradictory) model of VG. Specifically, VG asserts that differences in validity across studies are largely attributable to sampling error and “artifactual” sources of variance such as range restriction and measurement error (i.e., attenuation). Schmidt et al. (1976, p. 484) state that “our excessive faith in small-sample studies may account for much, if not all, of the variance in the phenomenon of ‘validity specificity’” and Schmidt and Hunter (1977, p. 219) claim that “if the variance in validity coefficients across situations for job-test combinations is due to statistical artifacts [i.e., study artifacts], then obviously the doctrine of SS is false and validities are generalizable” (p. 219).

VG and its related ideas and analyses have since become common in organizational literature. Aguinis et al. (2011) report that 83.5% of our published meta-analytic effect size (MAES) estimates rely on the Schmidt and Hunter procedure (with more than half adjusting MAES estimates based on statistical artifacts). However, despite its prevalence in organizational meta-analyses, VG has always been controversial. For example, James et al. (1986) argued that VG's primary logical argument is guilty of affirming the consequent.¹ James et al. (1992) demonstrated that variance in study context may be correlated with statistical artifacts, meaning that these sources of variance

¹Specifically, if validity is generalizable, variation in effect sizes will be explained by statistical artifacts. However, this does not logically imply that the ability to explain effect size variation with statistical artifacts means that validity is generalizable (as alternative explanations exist).

cannot be treated as independent or contradictory, as is assumed in VG. Other research demonstrated that statistical artifacts cannot be assumed to be independent from one another, as is also assumed in VG (Callender & Osburn, 1980; Köhler et al., 2015). VG has also been repeatedly challenged in court, with mixed results (Kleiman & Faley, 1985; Seymour, 1988).

Relevant meta-analytic considerations

Fixed versus random effects models

Early meta-analyses relied on a fixed-effect model, in which all studies share a common effect size and any observed differences between primary study effect sizes are due to sampling error (Borenstein et al., 2009). However, soon after the introduction of meta-analysis, Larry Hedges (1983, p. 388) developed an alternative random-effects model in which “characteristics of a study may influence the magnitude of its effect size.” Specifically, a random-effects model assumes that there is a distribution of true effect sizes that vary by study, as opposed to the common effect size assumed by a fixed-effect model (see Hedges & Olkin, 1985). Variance in this distribution can be attributed to factors beyond sampling error, including differences in samples, study designs, interventions, and measures. Accordingly, Hedges (1983) noted that interpreting the average effect size in a random-effects model is not meaningful in the absence of an estimate of variation.

The vast majority of meta-analyses conducted in organizational scholarship currently rely on random-effects models (87.5%, according to Aguinis et al., 2011). As such, the organizational literature seems to have come to a near consensus that differences in study context are meaningful, and we expect validity estimates to vary on the basis of these differences. At first glance, this may seem to refute the idea that validity is generalizable. However, it is important to consider that a fixed-effect model specifies that effect size variation is *only* attributable to sampling error. By accounting for measurement error (a function of measurement) and range restriction (a function of the sample), VG is intentionally accounting for artifactual sources of effect size variation attributable to study characteristics and is generally considered a random-effects model. Because the VG perspective seeks to explain effect size variation as a function of *artifactual* variance between primary studies, an evaluation of the VG perspective needs to focus on whether meaningful (i.e., non artifactual) sources of effect size variation also exist.

Heterogeneity and moderation

As Hedges (1983) specified, reliance on a random-effects model necessitates estimation of effect size variability. This estimate is referred to as heterogeneity, and meta-analysts often seek to explain this heterogeneity by proposing meta-analytic moderators that may account for variation between primary study effect sizes. Meta-analytic moderators pose a question directly relevant to the present discussion, as they attempt to determine whether effect sizes vary as a function of primary study characteristics.

A fixed-effect model does not allow for the existence of heterogeneity. VG allows for heterogeneity but posits that variance in primary study effect sizes is explicable as a function of artifactual sources of variance. SS assumes that effect sizes vary in meaningful ways. So which explanation is correct? A number of reviews have addressed this topic. Cortina (2003, p. 426) concludes that “a relatively small percentage [21.7%] of variance is typically attributable to artifacts, and considerable variability remains after correction of variance for artifacts.” Carlson and Ji (2011) report that 73% of SD_{ρ} values exceed .05, which limits the generalizability of an MAES. Finally, Tett et al. (2017) estimate that accounting for meta-analytic moderators reduces heterogeneity by 73.1%. In summary, literature examining the nature of heterogeneity seems to heavily favor SS. Artifact-based, between-study variation is present, suggesting that the VG perspective has a role to play. However, the literature examining meta-analytic heterogeneity and moderation suggests that primary study contexts also vary in meaningful ways.

The focal article's findings on selection

Consistent with prior perspectives on heterogeneity (Hedges, 1983; see also DeSimone et al., 2019; Murphy, 2017), the focal article acknowledges the importance of attending to variability in primary study validity estimates. Additionally, the focal article (p. 8) contends that the best predictors of job performance are those “specific to individual jobs, such as structured interviews, job knowledge tests, work sample tests, and empirically keyed biodata.”² The focal article also suggests that custom-designed selection systems are preferable to off-the-shelf systems, further emphasizing the importance of considering job-specific factors in selection. The idea that validity generalizes across contexts seems incompatible with the focal article's suggestion that the best predictor for a given job depends heavily on the job in question.

Recommendations 4 to 7 from the focal article suggest reliance on local (i.e., primary-study specific) or contextually appropriate adjustments for attenuation and range restriction. It seems obvious that meta-analysts who want to “correct” (i.e., adjust) for attenuation and range restriction should aim to do so on the basis of “correct” (i.e., maximally accurate) estimates of study artifacts. As such, we particularly appreciate Oh et al's (in press) for demonstration that range restriction is still present in concurrent validation studies (even if it is less extreme). Searching for appropriate estimates of range restriction in these studies is more appropriate than assuming that range restriction is absent. We hope articles like these will encourage more primary studies to report information relevant to psychometric adjustments. In doing so, meta-analysts can compute more accurate MAES estimates, and methodologists can better evaluate how variance in study artifacts varies across contexts and influences reported effect sizes.

Can validity generalization and situational specificity coexist?

To summarize, organizational scholarship has largely adopted random-effects models for meta-analysis, acknowledging effect size variance across contexts. The majority of meta-analyses published in our field demonstrate heterogeneity beyond what can be accounted for by study artifacts and attempt to account for effect size variance using meaningful substantive and methodological moderators. The best predictors of job performance vary by job. This evidence seems to point to SS as a plausible model, as validity estimates vary meaningfully as a function of context. If forced to choose between VG and SS, it would be reasonable to conclude in favor of the latter. However, we believe this “choice” is overly simplistic, as both perspectives have valuable insights to offer. Despite strong conceptual and empirical support for SS, we believe that VG and its underlying ideas still play an important role in organizational scholarship.

The VG perspective introduced psychometric adjustments for study artifacts to the conduct of meta-analysis. It would be difficult to argue that these adjustments have not had a strong impact on the conduct of meta-analysis in our field, a point furthered by the focal article's aim of improving the accuracy of range restriction estimates. Adjustments for attenuation and range restriction are useful when estimating the relationships between constructs but are not intended to indicate the relationships that individual primary studies or local validation efforts will observe when analyzing their own data (Biddle & Nooren, 2006; DeSimone, 2014). “Operational” validity estimates do not adjust validities for predictor unreliability because “in actual selection use, one must use the test as it exists—unreliability and all” (Schmidt et al., 1976, p. 474). In reality, organizations are stuck with not only their unreliable predictor measures but also their unreliable criterion measures and their observed ranges/variances. Practitioners should be wary of psychometric adjustments and omnibus MAES estimates, knowing that the validity they will most likely encounter in practice is closer to an unadjusted moderator-specific MAES computed using

²In fairness to Schmidt and Hunter (1998), these four job-specific predictors fare well in their analysis as well (see Table 1 on page 265, with operational validity estimates of .51, .48, .54, and .35, respectively, compared to .51 for general mental ability and .31 for conscientiousness).

primary studies that closely match their specific context. However, scholars are often concerned with a broader perspective, as our theoretical and empirical models are maximally useful when applied more generally. As such, VG is intuitively appealing to academicians concerned with broader, construct-level relationships. For these scholars, adjusted omnibus estimates can be informative, though we agree with the focal article that local adjustments are far more appropriate than global adjustments, as they do not assume that artifacts operate similarly across contexts.

A compromise between VG and SS may already exist. LeBreton et al. (2017) argue that VG and SS are not mutually exclusive, proposing a continuum of VG from “strong VG” (in which all effect size variation can be accounted for by sampling error and study artifacts) to “strong SS” (in which meaningful variation exists between primary studies). They also propose a middle ground in which meaningful variation exists, but some level of validity exists across contexts. Similarly, Cortina (2003) differentiated between the goals of transportability (established when the omnibus MAES is statistically different from zero) and parameter estimation (in which researchers attempt to accurately estimate effect sizes and account for heterogeneity using meta-analytic moderators). From these perspectives, it is possible for validity to generalize in direction but not magnitude. For example, we may know that integrity has a negative correlation with counterproductive work behaviors across (most) contexts but still acknowledge that this negative correlation is relatively strong in some situations and relatively weak in others. Tett et al. (2017) proposed the idea of a trade-off between precision of MAES estimates and generalizability of meta-analytic findings. More general statements such as “X is positively related to Y” may be more confidently generalizable across settings than more specific statements such as “the relationship between X and Y is between .40 and .50.”

Conclusion

In conclusion, like most academic arguments, the truth likely lies somewhere in the middle. VG was originally proposed as a counterargument to the notion that validity estimates were specific to particular contexts, fulfilling an academic desire to generalize research findings and maximize their utility. But VG, at least in its strongest form, is indefensible when confronted with repeated demonstrations of effect size variability attributable to meaningful contextual factors and evidence (such as that presented in the focal article) that the best predictors of job performance vary by job. Nevertheless, we argue that there is no need to throw the baby out with the bathwater—VG and SS can coexist. In revisiting SS, we conclude that it is a useful model and that contextual factors are well worth considering when discussing validity. With reference to meta-analysis specifically, reporting and/or interpreting a single MAES estimate is not advisable, especially when the meta-analysis reports high levels of heterogeneity and/or empirically supported meta-analytic moderation. However, our conclusion does not obviate the utility of VG. Broad statements about validity may generalize across settings without the need for specific estimates of validity to generalize. Meta-analysts can establish that an effect exists across many settings while exploring how the magnitude of that effect changes according to various jobs, contexts, and situations. As the focal article suggests, psychometric adjustments can account for situational differences by conducting these analyses in each primary study instead of relying on global adjustment techniques. Readers can interpret moderator-specific MAES estimates relevant to their interests or present contexts instead of interpreting omnibus MAES estimates. In doing so, researchers can appreciate the benefits of meta-analysis (including some aspects of VG) while still acknowledging that there are meaningful situational characteristics that influence the magnitude of a validity estimate. VG has plenty of methodological benefits to offer, but its contrast with SS is a false dichotomy. It is possible to reap these benefits while simultaneously acknowledging that heterogeneity in effect sizes is not entirely (or even primarily) due to study artifacts. It is not time to abandon validity generalization, but yes, it is time to revisit situational specificity.

References

- Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M.** (2011). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*, *37*, 5–38.
- Albright, L. E., Glennon, J. R., and Smith, W. J.** (1963). The uses of psychological tests in industry. Howard Allen.
- Biddle, D. A., & Nooren, P. M.** (2006). Validity generalization vs. Title VII: Can employers successfully defend tests without conducting local validation studies? *Labor Law Journal*, *57*, 216–237.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R.** (2009). *Introduction to meta-analysis*. Wiley.
- Callender, J. C. & Osburn, H. G.** (1980). Development and test of a new model for validity generalization. *Journal of Applied Psychology*, *65*, 543–558.
- Carlson, K. D., & Ji, F. X.** (2011). Citing and building on meta-analytic findings: A review and recommendations. *Organizational Research Methods*, *14*, 696–717.
- Cortina, J. M.** (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods*, *6*, 415–439.
- DeSimone, J. A.** (2014). When it's incorrect to correct: A brief history and cautionary note. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *7*, 527–531.
- DeSimone, J. A., Köhler, T., & Schoen, J. L.** (2019). If it were only that easy: The use of meta-analytic research by organizational scholars. *Organizational Research Methods*, *22*, 867–891.
- Ghiselli, E. E.** (1966). *The validity of occupational aptitude tests*. Wiley.
- Guion, R. M.** (1965). *Personnel testing*. McGraw-Hill.
- Hedges, L. V.** (1983). A random effects model for effect sizes. *Psychological Bulletin*, *93*, 388–395.
- Hedges, L. V. & Olkin, I.** (1985). *Statistical models for meta-analysis*. Academic Press.
- James, L. R., Demaree, R. G., & Mulaik, S. A.** (1986). A note on validity generalization procedures. *Journal of Applied Psychology*, *71*, 440–450.
- James, L. R., Demaree, R. G., Mulaik, S. A., & Ladd, R. T.** (1992). Validity generalization in the context of situational models. *Journal of Applied Psychology*, *77*, 3–14.
- Kleiman, L. S., & Faley, R. H.** (1985). The implications of professional and legal guidelines for court decisions involving criterion-related validity: A review and analysis. *Personnel Psychology*, *38*, 803–833.
- Köhler, T., Cortina, J. M., Kurtessis, J. N., & Gözl, M.** (2015). Are we correcting correctly? Interdependence of reliabilities in meta-analysis. *Organizational Research Methods*, *18*, 355–428.
- Lawshe, C. H.** (1948). *Principles of personnel testing*. McGraw-Hill.
- LeBreton, J. M., Schoen, J. L., & James, L. R.** (2017). Situational specificity, validity generalization, and the future of psychometric meta-analysis. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (2nd ed.) (pp. 93–114). Routledge.
- Murphy, K. R.** (2017). What inferences can and cannot be made on the basis of meta-analysis? *Human Resource Management Review*, *27*, 193–200.
- Oh, I., Le, H., & Roth, P. L.** (in press). Revisiting Sackett et al.'s (2022) recommendation against correcting for range restriction in concurrent validation studies. *Journal of Applied Psychology*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4308528
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F.** (2023). Revisiting the design of selection systems in light of new findings regarding the validity of widely used predictors. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *16*(3), 283–300.
- Schmidt, F. L., & Hunter, J. E.** (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*, 529–540.
- 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*, 262–274
- Schmidt, F. L., Hunter, J. E., & Urry, V. W.** (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology*, *61*, 473–485.
- Seymour, R. T.** (1988). Why plaintiff's council challenge tests, and how they can successfully challenge the theory of "validity generalization". *Journal of Vocational Behavior*, *33*, 331–364.
- Tett, R. P., Hundley, N. A., & Christiansen, N. D.** (2017). Meta-analysis and the myth of generalizability. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *10*, 421–456.

Cite this article: DeSimone, J. A. & Fezzey, T. N. A. (2023). Is it also time to revisit situational specificity? *Industrial and Organizational Psychology* *16*, 317–321. <https://doi.org/10.1017/iop.2023.40>