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. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage.
Schmidt, F. L., Oh, I.-S., & Hayes, T. (2009). Fixed versus random effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62*, 97–128.
Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA: Hogrefe & Huber.
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*(3), 421–456.

Validity Generalization as a Continuum

Ernest H. O'Boyle
University of Iowa

Tett, Hundley, and Christiansen (2017) make a compelling case against meta-analyses that focus on mean effect sizes (e.g., r_{xy} and ρ) while largely disregarding the precision of the estimate and true score variance. This is a reasonable point, but meta-analyses that myopically focus on mean effects at the expense of variance are not examples of validity generalization (VG)—they are examples of bad meta-analyses. VG and situational specificity (SS) fall along a continuum, and claims about generalization are confined to the research question and the type of generalization one is seeking (e.g., directional generalization, magnitude generalization). What Tett et al. (2017) successfully debunk is an extreme position along the generalization continuum significantly beyond the tenets of VG that few, if any, in the research community hold. The position they argue against is essentially a fixed-effects assumption, which runs counter to VG. Describing VG in this way is akin to describing SS as a position that completely ignores sampling error and treats every between-sample difference in effect size as true score variance. Both are strawmen that were knocked down decades ago (Schmidt et al., 1985). There is great value in debating whether a researcher should or can argue for generalization, but this debate must start with (a) an accurate portrayal of VG, (b) a discussion of different forms of generalization, and (c) the costs of trying to establish universal thresholds for VG.

First, true score variance is not the enemy of VG. Rather, it is at the heart of it. By partitioning observed variance into that which can be can considered

Ernest H. O'Boyle, Tippie College of Business, University of Iowa.
Correspondence concerning this article should be addressed to Ernest H. O'Boyle, Tippie College of Business, University of Iowa, 21 E. Market St., Iowa City, IA 52242. E-mail: oboyleeh@gmail.com

https://doi.org/10.1017/iop.2017.46 Published online by Cambridge University Press

artifact and that which can be considered true variance, VG proponents seek the *extent* to which an effect can be generalized, not *whether* it is generalizable. There are times where a dichotomous judgment may be based on the extent of generalization (e.g., whether a selection tool is legally defensible or not), but this does not negate the underlying continuous nature of VG.

Second, VG has different applications, and there is no one threshold of VG, just as there is no one true form of SS. The appropriate point along the continuum depends on the degree of generalization one is seeking. If the primary interest is in effect direction, then a wide credibility interval (CV) that nonetheless does not include zero offers some support of directional generalization. If the interest is in magnitude, then a CV that spans .20 to .40 may be sufficient to claim magnitude generalization. In a similar vein, the same CV width can lead to different conclusions about the extent of generalization. An 80% CV width of .05 for the relation between a well-validated scale and some objective, high-stakes outcome may rightly send the meta-analyst doggedly seeking out those subpopulations and moderators that explain this unaccounted for variance. On the other hand, the same CV width may be considered quite generalizable if, for example, the research question was one of construct redundancy (Le, Schmidt, Harter, & Lauver, 2010), and the CV spanned .94 to .99. Seeking out those conditions where the two constructs were completely redundant as opposed to nearly redundant may be of little interest.

Third, by artificially dichotomizing meta-analytic results as either supporting VG or supporting SS, Tett et al. (2017) find themselves in the ironic position of defending situational specificity by way of broad generalizations (i.e., cut-off values for SDρ and SE-r_{xy}). One problem with universal cut-off values for VG is that they run counter to the different forms and applications of VG, but these threshold values are problematic for other reasons. To begin with, by suggesting cutoff values in any form, Tett et al. make an implicit assumption that there is a universal standard for calculating SDρ and SE-r_{xy}. Yet, for psychometric meta-analysis, both SDρ and SE-r_{xy} differ in calculation depending on which statistical artifacts are corrected for and the type of artifact correction employed. However, the larger issue is the specific cutoff values proposed by Tett et al. Their basis on Oswald and McCloy (2003) is tenuous. Oswald and McCloy (2003) issued a guarded statement in their book chapter that .025, .075, and .125 could be thought of as small, medium, and large values of SDρ. However, in the very same paragraph they state, "Of course, different values could be set as standards" (Oswald and McCloy, 2003, p. 325). Their point was that "a large amount of true score variance should not be overlooked" (p. 325). That is a far cry from "We suggest that Oswald and McCloy's SD(rho) $<$ .025 offers a reasonable, albeit rigorous, standard [for VG justification]" (Tett et al., 2017, p. 446). Tett

https://doi.org/10.1017/iop.2017.46 Published online by Cambridge University Press

et al. then "tentatively" extend this .025 cut-off value to SE-r_{xy}. Tentatively or not, an inadequately justified cutoff applied broadly is how myths and urban legends develop (Lance, Butts, & Michels, 2006). Whether it is the .70 threshold for reliability, the .90 cut-off for the Comparative Fit Index, or even the 75% rule for psychometric meta-analysis, no amount of debunking, clarifications, or outright recanting can fully put the genie back in the bottle. That other somewhat arbitrary thresholds for VG exist is not justification for additional thresholds that attempt the impossible task of "generalizing" generalization.

Further, offering threshold values that allow authors to make a definitive statement of VG encourages questionable research and reporting practices (O'Boyle, Banks, & Gonzalez-Mulé, 2017). Specifically, if a belief develops that the likelihood of publication will be improved by claiming VG based on an 80% CV less than ±.032 and a 95% CI less than ±.05, then true score variance *will* become the enemy of VG. Just as we have seen countless times with null hypothesis significance testing (Cortina & Landis, 2001; Orlitsky, 2012) and more recently with structural equation modeling (Cortina, Green, Keeler, & Vandenberg, 2017), the introduction of even well-intentioned cut-offs will encourage inappropriate decisions about outliers, inclusion criteria, and reporting standards.

Ultimately, attempts to pit VG against SS oversimplify both positions by dichotomizing generalization. This leads to mischaracterizing VG as a narrowly defined, rigid set of standards where rho is the cynosure of all eyes and all variance is simply statistical artifact in need of correction. (It is worth noting that VG advocates are just as prone to mischaracterizing and oversimplifying SS.) By viewing validity generalization along a continuum, meta-analysts are able to make informed decisions about the extent to which an effect generalizes and what the implication of that generalization means for that specific research question.

References

Cortina, J. M., Green, J. P., Keeler, K. R., & Vandenberg, R. J. (2017). Degrees of freedom in SEM: Are we testing the models that we claim to test? *Organizational Research Methods, 20*(3), 350–378.

Cortina, J. M., & Landis, R. S. (2011). The earth is not round (p =. 00). *Organizational Research Methods, 14*(2), 332–349.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*(2), 202–220.

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 and Human Decision Processes, 112*(2), 112–125.

O'Boyle, E. H., Banks, G.C., & Gonzalez-Mulé, E. (2017). The Chrysalis Effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management, 43*, 376–399.

Orlitzky, M. (2012). How can significance tests be deinstitutionalized? *Organizational Research Methods, 15*(2), 199–228.

https://doi.org/10.1017/iop.2017.46 Published online by Cambridge University Press

Oswald, F. L., & McCloy, R. A. (2003). Meta-analysis and the art of the average. In K. Murphy (Ed.), *Validity generalization: A critical review* (pp. 311–338). Mahwah, NJ: Erlbaum.

Schmidt, F. L., Hunter, J. E., Pearlman, K., Hirsh, H. R., Sackett, p. R., Schmitt, N., … Sedeck, S. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology, 38*(4), 697–798.

Tett, R. P., Hundley, N., & Christiansen, N. D. (2017). Meta-analysis and the myth of generalizability. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 10*(3), 421–456.

A Failed Challenge to Validity Generalization: Addressing a Fundamental Misunderstanding of the Nature of VG

Frank L. Schmidt
University of Iowa

Chockalingam Viswesvaran
Florida International University

Deniz S. Ones
University of Minnesota

Huy Le
University of Texas at San Antonio

The lengthy and complex focal article by Tett, Hundley, and Christiansen (2017) is based on a fundamental misunderstanding of the nature of validity generalization (VG): It is based on the assumption that what is generalized in VG is the estimated value of mean rho ($\bar{\rho}$). This erroneous assumption is stated repeatedly throughout the article. A conclusion of validity generalization does not imply that $\bar{\rho}$ is identical across all situations. If VG is present, most, if not all, validities in the validity distribution are positive and useful even if there is some variation in that distribution. What is generalized is the entire distribution of rho ($\bar{\rho}$), not just the estimated $\bar{\rho}$ or any other specific value of validity included in the distribution. This distribution is described by its mean ($\bar{\rho}$) and standard deviation (SD_ρ). A helpful concept based

Frank L. Schmidt, University of Iowa; Chockalingam Viswesvaran, Florida International University; Deniz S. Ones, University of Minnesota; Huy Le, University of Texas at San Antonio.

The order of authorship following the lead author is by seniority. We thank Philip Roth and In-Sue Oh for suggestions and comments on an earlier draft of this manuscript.

Correspondence concerning this article should be addressed to Chockalingam Viswesvaran, Florida International University, Department of Psychology, University Park, Miami, FL 33199. E-mail: Vish@fiu.edu; or Deniz S. Ones, University of Minnesota, Department of Psychology, 75 East River Road, Minneapolis, MN 55455. E-mail: Deniz.S.Ones-1@tc.umn.edu

https://doi.org/10.1017/iop.2017.46 Published online by Cambridge University Press