

- Wilmot, M. P. (2017). *Personality and its impacts across the behavioral sciences: A quantitative review of meta-analytic findings*. Doctoral dissertation. University of Minnesota, Minneapolis, MN.
- Zimmerman, R. D. (2008). Understanding the impact of personality traits on individuals' turnover decisions: A meta-analytic path model. *Personnel Psychology*, 61(2), 309–348. <https://doi.org/10.1111/j.1744-6570.2008.00115.x>

## On the Mystery (or Myth) of Challenging Principles and Methods of Validity Generalization (VG) Based on Fragmentary Knowledge and Improper or Outdated Practices of VG

In-Sue Oh  
*Temple University*

Philip L. Roth  
*Clemson University*  
*Baruch College, CUNY*

In their focal article, Tett, Hundley, and Christiansen (2017) stated in multiple places that if there are good reasons to expect moderating effect(s), the application of an overall validity generalization (VG) analysis (meta-analysis) is “moot,” “irrelevant,” “minimally useful,” and “a misrepresentation of the data.” They used multiple examples and, in particular, a hypothetical example about the relationship between agreeableness and job performance. Four noteworthy problems with the above statements, other similar statements elsewhere in Tett et al.'s article, and their underlying assumptions are discussed below along with alternative perspectives.

### **VG as a Method Should Be Distinguished From VG as a Practice**

Throughout the article, Tett et al. (2017) did not make it clear whether they intended to challenge either a statistical method of VG or some

In-Sue Oh, Department of Human Resource Management, Fox School of Business, Temple University; Philip L. Roth, Department of Management, College of Business, Clemson University.

We thank Chris Berry, Ernest O'Boyle, and Frank Schmidt for their helpful comments on an earlier version of this commentary.

Correspondence concerning this article should be addressed to In-Sue Oh, Department of Human Resource Management, Fox School of Business, Temple University, 1801 Liacouras Walk, Philadelphia, PA 19122-6083. E-mail: [insue.oh@temple.edu](mailto:insue.oh@temple.edu); and Philip L. Roth, Department of Management, College of Business, Clemson University, Clemson, SC 29634-1305. E-mail: [ROTHP@clemson.edu](mailto:ROTHP@clemson.edu)

improper or outdated practices of VG, but it appears that they challenged the accuracy and usefulness of VG as a method based on improper or outdated applications of VG as noted below. This confusion is problematic because VG (originally developed by Schmidt and Hunter [1977]) is a well-established statistical method/tool with statistical accuracy and efficiency verified in many articles (e.g., Field, 2001) and books by external evaluators (e.g., Schulze, 2004). This does not necessarily mean that VG cannot be misused and abused in its applications like other well-established statistical analysis methods (e.g., regression, hierarchical linear modeling, structural equation modeling). Thus, there is a need to clearly distinguish VG as a method from VG as a practice. Like some recent challenges to VG and meta-analysis (e.g., James & McIntyre, 2010; Muchinsky & Raines, 2013), Tett et al. also failed to make this distinction and, thus, confused the readers of their article as discussed in some detail in this commentary.

### **VG (or Lack of It) Should Be Treated as a Matter of Degree, Not Dichotomy**

Tett et al. (2017) stated that VG is determined using some rules such as the 75% rule and/or the 90% CV rule.<sup>1</sup> For example, they stated that “the 75% rule (Schmidt & Hunter, 1977) holds that if %VE > 75%, situational generalizability of mean rho may be inferred” (p. 425) and that “VG dichotomizes the continuum of correlation strength (James & McIntyre, 2010); in terms of the “90% [CV] > 0 rule, either the 90% CV falls above 0, conferring VG, or it does not, failing to confer VG” (p. 426; bracket added for clarity). However, more recent books and articles about VG have made it clear that when examining VG analysis results, we need to regard VG (and situational specificity) *as a matter of degree, not dichotomy*. In fact, for this reason, Hunter and Schmidt dropped the 75% “rule of thumb,” which was originally included in the first edition of their meta-analysis book (Hunter & Schmidt, 1990), from the second and third editions of the book (Hunter & Schmidt, 2004; Schmidt & Hunter, 2015). Note that like many other well-established statistical analysis methods, meta-analysis (VG as a method) is a constantly evolving research synthesis tool, and meta-analysts should be aware of major advancements and refinements of the method (Cortina, Aguinis, & DeShon, 2017; Schmidt, 2008). Thus, VG results should also be interpreted as a matter of degree, not a matter of dichotomy (VG or not) in order not to be subject to the same fallacy created by null hypothesis significance testing (NHST; significant or not). Meta-analysis should be practiced by scientists, not judges (with a “yes” or “no” switch in their heads). In the same vein:

<sup>1</sup> Tett et al. (2017) termed them as “rules” in many places of their article, but they are just *rules of thumb*, not absolute rules.

It is important to note, however, that validity generalization can be justified in many cases even if the remaining variance [i.e.,  $SD(\rho)$ ] is not zero. That is, validity generalization can be justified in many cases in which the hypothesis of situational specificity cannot be definitively rejected. (Pearlman, Schmidt, & Hunter, 1980, p. 376; bracket added)

Finally, to unstrap the straitjacket of NHST or such dichotomous heuristics, the degree of VG should be carefully gauged by triangulating all available meta-analytic results and, if possible, relevant prior VG results as discussed next (e.g., Pearlman et al., 1980; Salgado, Anderson, Moscoso, Bertua, & Fruyt, 2003).

### **The Degree of VG Should Be Gauged by Triangulating All Available Meta-Analytic Results**

Tett et al. (2017) stated in multiple places of their article that if there are some good reasons to expect some moderating effect(s) results from an overall VG analysis (meta-analysis) are “moot,” “irrelevant,” “minimally useful,” and “a misrepresentation of the data” using multiple examples. They developed a hypothetical, and atypical, example of an overall VG analysis about the bidirectional relationship between agreeableness and job performance across different samples (e.g., positive relationships for jobs “where caring for others is especially valued” and negative relationships for jobs where being “tough skinned is favored”). Before commenting more on the specific hypothetical example, we would like to note two things. First, this example by Tett et al. does not represent a problematic meta-analysis practice that mixes apples, oranges, and pears in meta-analysis (e.g., mixing many different personality traits in the same meta-analysis and concluding that personality does not matter in predicting performance because of the low mean  $\rho$  and a huge amount of variability [ $SD(\rho)$ ] across input validities; Cortina, 2003). Second, the vast majority of moderator analyses in meta-analysis/VG do not appear to result in negative mean validities for predictors such as cognitive ability tests (Hunter & Hunter, 1984; Salgado et al., 2003), conscientiousness (Barrick & Mount, 1991), work sample tests (Roth, Bobko, & McFarland, 2005), assessment centers (Gaugler, Rosenthal, Thornton, & Bentson, 1987), grade point average (GPA; Roth, BeVier, Switzer, & Schippmann, 1996), or employment interviews (Huffcutt & Arthur, 1994). Instead, such analyses typically result in varying degrees of positive relationships, suggesting there could be useful validities at multiple values or levels of moderators (e.g., positive validities for GPA across multiple conditions). In a very real sense, there is an argument that the data support validity across multiple levels of a moderator (e.g., see Hunter & Hunter’s [1984] work on useful levels of validity across levels of job complexity) and that the judgment of validity does generalize across those moderator levels.

Now to Tett et al.'s (2017) example: Although the percentage of jobs in which a low level of agreeableness is valued is much smaller than jobs in which a high level of agreeableness is valued, let us conservatively assume that we have six validation studies on the relationship between agreeableness and job performance with a sample size of 150 each (a reasonable sample size found in many validation studies), and that observed validities are .09 (a reasonable value for this trait; Barrick & Mount, 1991) in three studies and observed validities are  $-.09$  in the remaining three studies. Assuming no measurement error and no range restriction in all six input studies (just to make the example easy to understand), overall VG analysis results will show that the mean  $\rho$  is .00 and  $SD(\rho)$  is .04. The 80% credibility interval ranges from  $-.05$  to .05, and the 95% confidence interval ranges from  $-.07$  to .07; the percentage of variance due to artifacts (%VE; sampling error alone in this case) is 83%. If meta-analysts followed the VG practices discussed in Tett et al.'s article, they would be confused given that the 90% CV "rule of thumb" would suggest that VG is not present, whereas the 75% "rule of thumb" would suggest that VG is present. Again, VG should not be determined in a dichotomous manner using only part of the meta-analytic results, although improper and outdated practices of VG may have determined VG in this fashion. Instead, the aforementioned VG results suggest that we cannot rule out the possibility that the sign of the input validities is mostly artifactual due to sampling error, not due to occupational differences given the reliability for the validity distribution (vector) is very low at .17 ( $= 1 - .83$ ) or the correlation between the observed validities and sampling error is .91 ( $= \text{SQRT of } .83$ ). That is, we simply cannot completely trust any of the observed validities in the distribution (at face value) because the magnitude and sign of those validities are mostly due to sampling error.<sup>2</sup> Obviously, this is a more parsimonious explanation for the apparent variation in the entire validity distribution. Furthermore, it is well known that %VE is a percent-based relative index (in this case,  $.83 = [.0067 / .0081]$ ) and, thus, should be triangulated by carefully considering its components,<sup>3</sup> as well as all other meta-analytic results such as  $k$ , total  $N$ ,  $SD(\rho)$ ,

<sup>2</sup> Schmidt and Hunter (2015, pp. 107–110) noted that  $1 - \%VE$  represents the reliability of the validity distribution (vector) because  $[1 - \%VE]$  equals the ratio of  $\text{Var}(\rho)$  to  $\text{Var}(r)$ ; if  $[1 - \%VE]$  (the reliability of the validity distribution) is close to zero, the observed variance across observed validities is mostly artifactual). Schmidt and Hunter further noted that the square root of %VE represents the correlation between observed validities and sampling error.

<sup>3</sup> A 10% variance due to artifacts could be a small or huge amount depending on the size of the denominator (e.g.,  $1/10$  or  $.001/.0001$  or  $.000001/.0000001$ ).

SE(mean  $\rho$ )<sup>4</sup>), and their extensions (the 80% credibility and 95% confidence intervals).

### **The Mean Rho From an Overall VG Analysis Is Useful as Long as the Analysis Is Properly Conducted**

Tett et al. (2017) claimed that the mean rho estimated from an overall VG analysis is “moot” or “irrelevant” if moderating effects are expected with good reasons. Simply put, the overall mean rho estimated via meta-analysis, regardless of its magnitude, is the best estimate for mean rho of the grand population and thus represents the grand population (that may subsume subpopulations) more accurately than any other single value. The estimated subgroup mean rho via meta-analysis is the best estimate for mean rho of the subpopulation and thus accurately represents the subpopulation. For clarity, we note that each subpopulation may have some known or unknown moderators (i.e., next lower-level subpopulations), and thus mean rho, not rho, is used here according to the primary principle of the random-effects model (Schmidt, Oh, & Hayes, 2009). Of course, not surprisingly, the overall mean rho estimate does not more accurately represent subpopulations (i.e., some subsets/subgroups of the entire validity distribution) than corresponding subgroup mean rho estimates, and likewise, any of the subgroup mean rho estimates do not more accurately represent the grand population (the entire validity distribution) than the overall mean rho estimate. Accordingly, Tett et al.’s claim that mean rho estimated from an overall VG analysis misrepresents the entire validity distribution if some moderating effects are expected is inappropriate as long as the VG analysis is *properly* conducted. Of course, if an overall VG analysis was conducted improperly by mixing different constructs (e.g., general mental ability, proactive personality) or different methods (e.g., work sample tests, employment interviews, situational judgment tests) as predictors of job performance, the mean rho estimated from the overall analysis would be almost moot and potentially meaningless (Cortina, 2003). As noted above, mean rho should not be the sole focus of VG.

### **Conclusion**

In conclusion, we want to draw the readers’ attention to the fact that VG is a well-established statistical method/tool that can be properly used, misused, and abused like other well-established statistical analysis methods. Like the results of many other statistical analysis methods, VG analysis results should

<sup>4</sup> Tett et al. used SE(r) in multiple places of their article, but it should read SE(mean r) or SE(mean rho), as meta-analysts desire to estimate the error band (confidence interval) around the mean r or mean rho, not an individual r or an individual rho in meta-analysis.

be interpreted as a matter of degree, not as a matter of dichotomy (VG or not), in order to be scientifically useful and not to be subject to the same fallacy created by NHST (which ironically VG was designed to address). In addition, VG analysis results should also be interpreted while considering and triangulating all available meta-analytic results, not just mean rho and/or SD(rho). Given that VG is a constantly evolving statistical method like many other statistical methods, meta-analysts should not stick to outdated practices and methods but keep abreast of important advancements and refinements in both VG practices and methods. Finally, there is a need to distinguish VG as a method from VG as a practice, and hence, improper or outdated VG practices should not be a basis for challenging VG as a state-of-the-art method.

## References

- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.
- Cortina, J. M. (2003). Apples and oranges (and pears, oh my!): The search for moderators in meta-analysis. *Organizational Research Methods, 6*, 415–439.
- Cortina, J. M., Aguinis, H., & DeShon, R. P. (2017). Twilight of dawn or of evening? A century of research methods in the *Journal of Applied Psychology, 102*, 274–290.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed-and random-effects methods. *Psychological Methods, 6*, 161–180.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*(3), 493–511.
- Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*, 184–190.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72–98.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings* (1st ed.). Thousand Oaks, CA: Sage.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting for error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- James, L. R., & McIntyre, H. H. (2010). Situational specificity and validity generalization. In J. L. Farr & N. T. Tippins (eds.), *Handbook of employee selection* (pp. 909–920). New York: Routledge.
- Muchinsky, P. M., & Raines, J. M. (2013). The overgeneralized validity of validity generalization. *Journal of Organizational Behavior, 34*, 1057–1060.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology, 65*, 373–406.
- Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology, 58*(4), 1009–1037.
- Roth, P. L., Bevier, C. A., Switzer, F. S., & Schippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology, 81*, 548–556.
- Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European Community meta-analysis. *Personnel Psychology, 56*, 573–605.
- Schmidt, F. L. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods, 11*, 96–113.

- 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  $\rho$ ) 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 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](mailto:oboyleeh@gmail.com)