



Contents lists available at SciVerse ScienceDirect

## Organizational Behavior and Human Decision Processes

journal homepage: [www.elsevier.com/locate/obhdp](http://www.elsevier.com/locate/obhdp)Methods for second order meta-analysis and illustrative applications <sup>☆</sup>Frank L. Schmidt <sup>a,\*,1</sup>, In-Sue Oh <sup>b,\*,1</sup><sup>a</sup> Department of Management and Organizations, Tippie College of Business, University of Iowa, Iowa City, IA 52242, United States<sup>b</sup> Department of Human Resource Management, Fox School of Business, Temple University, Philadelphia, PA 19122, United States

## ARTICLE INFO

## Article history:

Received 13 September 2012

Accepted 2 March 2013

Accepted by Xiao-Ping Chen

## Keywords:

Meta-analysis

Second order meta-analysis

Research synthesis

Sampling error

## ABSTRACT

This paper presents methods for second order meta-analysis along with several illustrative applications. A second order meta-analysis is a meta-analysis of a number of statistically independent and methodologically comparable first order meta-analyses examining ostensibly the same relationship in different contexts. First order meta-analysis greatly reduces sampling error variance but does not eliminate it. The residual sampling error is called second order sampling error. The purpose of a second order meta-analysis is to estimate the proportion of the variance in mean meta-analytic effect sizes across multiple first order meta-analyses attributable to second order sampling error and to use this information to improve accuracy of estimation for each first order meta-analytic estimate. We present equations and methods based on the random effects model for second order meta-analysis for three situations and three empirical applications of second order meta-analysis to illustrate the potential value of these methods to the pursuit of cumulative knowledge.

© 2013 Elsevier Inc. All rights reserved.

## Introduction

By integrating and synthesizing effect sizes across primary studies measuring ostensibly the same relation, a first order meta-analysis provides a mean effect size that is more accurate than any effect size available from the primary studies included in the meta-analysis (Hedges & Olkin, 1985; Hunter & Schmidt, 2004; McDaniel, 2007). One of the strengths of first order meta-analysis is its ability to reduce sampling error by synthesizing effect sizes estimates across multiple primary studies. However, given the total number of studies in any meta-analysis is less than infinite, this process does not reduce sampling error to zero; the remaining sampling error is called second order sampling error (Hunter & Schmidt, 2004, chap. 9). It is this second order sampling error that is the focus of second order meta-analysis, the methodology of conducting a meta-analysis of meta-analyses by synthesizing evidence from multiple meta-analyses. Second order meta-analysis is also known as overview of reviews, umbrella review, meta-meta-analysis, and meta-analysis of meta-analyses in other fields of the social sciences (e.g., Cooper & Koenka, 2012, p. 446).

A major goal of second order meta-analysis is to determine how much of the variance in mean effect sizes<sup>2</sup> across different first order meta-analyses of the same relation is due to second order sampling error (variance) and to use this information to improve estimates in individual meta-analyses.

Numerous meta-analyses have been conducted in various areas of industrial-organizational psychology and the related fields of Management, Organizational Behavior and Human resource Management. The same is true in other disciplines and research areas in psychology. For example, some research topics (e.g., the relation between personality and job performance; Oh, 2009) have been meta-analytically examined independently in multiple countries. As illustrated later, there are quite a few research areas in these literatures in which there are multiple independent first order meta-analyses on the same relationship. Hence there is need to integrate multiple meta-analyses conducted to estimate ostensibly the same relation by means of second order meta-analysis. Further, this need

<sup>2</sup> A mean effect size (e.g.,  $\hat{\rho}$ ) represents a first order meta-analytic estimate of effect sizes across statistically independent primary studies/samples computed using a meta-analytic methods based on the random-effects model; if the true or non-artifactual variance ( $\hat{\sigma}_{\rho}^2$ ) turns out to be zero, a meta-analytic (mean) effect size computed using a method based on the random-effects model ( $\hat{\rho}$ ) reduces to (equals) a meta-analytic effect size ( $\hat{\rho}$ ) computed using the fixed-effect model. Because the random-effects model subsumes the fixed-effect model (and *not* vice versa) and is more generally accepted than the fixed-effect model (Borenstein, Hedges, Higgins, & Rothstein, 2009; Schmidt, Oh, & Hayes, 2009), we use the term mean effect size ( $\hat{\rho}$ ) to represent a meta-analytic estimate for a relation computed using a method based on the random-effects model. Thus, a second order mean effect size ( $\hat{\bar{\rho}}$ ) represents the mean of the mean (meta-analytic) effect sizes across statistically independent individual first order meta-analyses.

<sup>☆</sup> This study was partly supported by following awards given to In-Sue Oh: The Meredith P. Crawford Fellowship in I/O Psychology by Human Resources Research Organization; The Ballard and Seashore Dissertation Year Fellowship by the Graduate College, University of Iowa; and The Center for Asian and Pacific Studies (CAPS) Graduate Dissertation Grant Award by the CAPS, University of Iowa.

\* Corresponding authors.

E-mail addresses: [frank-schmidt@uiowa.edu](mailto:frank-schmidt@uiowa.edu) (F.L. Schmidt), [insue.oh@temple.edu](mailto:insue.oh@temple.edu) (I.-S. Oh).<sup>1</sup> These authors equally contributed to this paper.

will increase in the future as more multiple meta-analyses accumulate in the literature. Cooper and Koenka (2012) examined methods for integrating results across different meta-analyses and concluded that none of the currently available methods are really satisfactory (particularly in estimating the amount of true variance). If we decide to synthesize effect size across the first order meta-analyses, there are three options to choose from as suggested in Borenstein, Hedges, Higgins, and Rothstein (2009, pp. 184–186).

The first option is to conduct a full meta-analysis including all primary studies. This is an omnibus meta-analysis pooling all primary studies across all potential moderators followed by separate meta-analyses for different potential moderators. As stated in Borenstein et al. (2009, chap. 19, p. 186), “if the subgrouping is not of major importance, or if multiple different subgroupings of the studies are being considered, then [this] is the more logical choice”. However, this option is possible only if all primary studies and data used in each first order meta-analysis are available to the researcher(s); in practice, this is often not the case. For example, studies in some of the meta-analyses might be written in languages that the researchers do not understand. In addition, although this procedure produces the same grand mean estimate as second order meta-analysis, it does not allow one to estimate the variance (and the percentage of that variance) across subgroup meta-analyses that is (and is not) due to second order sampling error, because the method does not allow second order sampling error variance to be computed. This option is frequently advocated and therefore we later discuss in more detail the disadvantages of this option in comparison to a second order meta-analysis. That material is presented later in this paper after the necessary conceptual foundations for understanding second order meta-analysis have been presented. The two options described next are viable when first order meta-analytic estimates are the only available input.

The second option (Borenstein et al., 2009, chap. 19) is to combine (i.e., average) mean effect sizes across first order meta-analyses of interest while ignoring the between-meta-analysis variance. Some scholars have conducted a second order analysis of this sort on the relations between personality and job performance across multiple, minimally overlapping, prior first order meta-analyses (e.g., Barrick, Mount, & Judge, 2001). According to Cooper and Koenka (2012, p. 458), this is the most common way in which second order meta-analysis is conducted at present. However, as discussed later, this option does not allow estimation of the amount of true (i.e., non-artifactual) variance between meta-analyses means (useful in estimating the credibility intervals for the second order meta-analytic means). Nor does it allow estimation of the amount of observed variation across meta-analyses that is due to second order sampling error (useful in estimating the confidence intervals for the second order meta-analytic means).

The third option (Borenstein et al., 2009, chap. 19) is to combine mean effect sizes across meta-analyses of interest while modeling the between-meta-analysis variance. This is the best option when primary studies from all relevant first order meta-analyses are unavailable (a second order meta-analysis can be conducted using only first order meta-analytic estimates), and there is a need to estimate the between-meta-analysis variance (for example, when each meta-analysis represents a random sample of a different population [e.g., country, occupation, type of industry, military vs. civilian]). As noted by Cooper and Koenka (2012, p. 458), complete statistical methods necessary for conducting this type of second order meta-analysis have not yet been introduced. Hunter and Schmidt (2004, pp. 406–408) briefly discussed how to compute second order sampling error but did not fully explicate the statistical methods necessary for estimating the between-meta-analysis variance. Borenstein et al. (2009, chap. 19, p. 183) acknowledged

that second order meta-analysis needs to model this additional source of variability, but “[t]he mechanism for doing so is beyond the scope of an introductory book”.

The major goal of this study is to present statistical methods for second order meta-analysis modeling between-meta-analysis variation with several illustrative applications. Later in this paper, after the conceptual and methodological bases for second order meta-analysis have been presented, we consider and address several potential objections to this procedure. The methods of second order meta-analysis are a straightforward generalization of first order random effects (RE) meta-analysis methods to the meta-analytic analysis of the mean effect size estimates across multiple meta-analyses. The methods presented in this paper cannot be applied to fixed effects (FE) meta-analyses. In response to critiques of the FE model (e.g., Hedges, 1994; Hunter & Schmidt, 2000; Schmidt et al., 2009), the FE meta-analysis model is rarely used today. This is because FE models assume *a priori* that there can be no real variation in effects sizes across studies (i.e., no variation beyond sampling error variance), an unrealistic assumption (Hunter & Schmidt, 2000; Raudenbush, 1994, 2009; Schmidt et al., 2009). This assumption would substantially bias downward the estimate of second order sampling error in any second order meta-analysis. The methods presented in this paper are based on the Hunter–Schmidt meta-analysis methods, which include only RE models. The second order meta-analysis methods introduced below are based on “a fully random-effects model” because they assume random effects both within and across meta-analyses (Borenstein et al., 2009, chap. 19, p. 183). To explicate this, we first present a brief review of the basic equations of first order meta-analysis, using the correlation coefficient as the illustrative statistic. Analogous equations apply for the *d* statistic. In first order meta-analysis, effect sizes from primary studies/samples are weighted by the inverse of their sampling error or a close approximation thereof. The following discussion of first order meta-analysis assumes such study weights. See Schmidt, Oh, and Hayes (2009; see also Brannick, Yang, & Cafri, 2011) for an extended discussion of the weighting of primary studies in first order meta-analysis.

## Review of first order meta-analysis methods

### Simple (bare bones) first order meta-analysis

The simplest form of meta-analysis is one in which only sampling error is taken into account, referred to as bare bones meta-analysis (Hunter & Schmidt, 2004, chap. 3). In this version of meta-analysis, there is no correction for measurement error. Eqs. (1) and (2) are the fundamental equations for this form of meta-analysis:

$$S_r^2 = S_{\rho_{xy}}^2 + E(S_{e_{r_i}}^2), \quad (1)$$

where  $S_r^2$  is the weighted variance of the observed correlations ( $r_i$ ) across statistically independent primary studies,  $S_{e_{r_i}}^2$  is the sampling error variance for each correlation, and  $E(S_{e_{r_i}}^2)$  is the weighted average of these sampling error variances. Transposing, we obtain Eq. (2):

$$\hat{\Sigma}_{\rho_{xy}}^2 = S_r^2 - E(S_{e_{r_i}}^2). \quad (2)$$

The term on the left in Eq. (2) is the variance of the actual population study correlations, estimated as the (weighted) observed variance of the correlations minus the expected (weighted average of) sampling error variance as computed from the usual formula for the sampling error variance of correlations (Hunter & Schmidt, 2004, chap. 3). If the term on the left side of Eq. (2) is zero or very small, the final result is the (weighted) mean observed correlation ( $\bar{r}$ ) and its standard deviation corrected for sampling error alone ( $SD_{\rho_{xy}}$ ). [Note that in a FE model, this value is assumed by fiat to

be zero.] If the term on the left side of Eq. (2) is large, there may be a search for moderators, which may be conducted either by subgrouping studies by moderator values and conducting separate meta-analyses or by regressing study correlations onto hypothesized moderators (i.e., meta-regression) (Aguinis, Gottfredson, & Wright, 2011; Borenstein et al., 2009, chaps. 19–20).

#### First order meta-analysis correcting each effect size individually

Measurement error is present in all measures used in all research and it biases all estimates of relationships in research. As a result, it is important to include corrections for these biases in order to obtain unbiased estimates of relationships. In comparison with bare bones meta-analysis, a more complete and accurate form of meta-analysis (Hunter & Schmidt, 2004, chapter 3) is one in which each correlation is first corrected for the downward biases created by measurement error (and for range restriction and dichotomization, if present). The meta-analysis is then performed on these corrected correlations (symbolized  $r_c$ ). Eq. (3) is the basic equation for this form of meta-analysis.

$$\hat{S}_\rho^2 = S_{r_c}^2 - E(S_{e_{r_c}}^2), \quad (3)$$

where the term on the left side of the equation is the estimated variance of the actual (disattenuated) population correlations,  $S_{r_c}^2$  is the weighted variance of the correlations that have been corrected individually for measurement error,  $S_{e_{r_c}}^2$  is the sampling error variance for each corrected correlation, and  $E(S_{e_{r_c}}^2)$  is the weighted mean across the corrected correlations of these sampling error variance values. In this form of meta-analysis, each corrected correlation ( $r_{ci}$ ) is weighted by  $N_i \times r_{xxi} \times r_{yyi}$ ; that is, by the three-way product of sample size ( $N_i$ ), the reliability of the independent variable measure ( $r_{xxi}$ ), and the reliability of the dependent variable measure ( $r_{yyi}$ ), where  $i$  indicates the  $i$ th study. This 3-way product represents the inverse of the sampling error variance of a correlation corrected for measurement error (Hunter & Schmidt, 2004, chap. 3). If there is, in addition, a correction for range restriction, this study weight becomes a 4-way product.

The first term on the right side of Eq. (3) is the observed weighted variance of the correlations that have been individually corrected for measurement error; this figure is typically larger than the weighted variance of the uncorrected correlations, because the correction for measurement error increases the variance of the correlations. The second term on the right is the expected weighted sampling error variance of these correlations. This sampling error is larger than in the case of Eq. (2) because the corrections for measurement error increase the sampling error in each correlation (Hunter & Schmidt, 2004, chap. 3). Procedures for calculating this sampling error variance are given in Hunter and Schmidt (2004, pp. 206–207). The term on the left side of Eq. (3) ( $\hat{S}_\rho^2$ ) is the estimate of the variance of the disattenuated population correlations (i.e., the population parameter true score correlations). If this term is zero or very small, the final result of the meta-analysis is the (weighted) mean corrected correlation ( $\hat{\rho}$ ) and its standard deviation corrected for sampling error and measurement error ( $SD_\rho$ ). If the term on the left side of Eq. (3) ( $\hat{S}_\rho^2$ ) is large, there may be a search for moderators.

#### First order artifact distribution meta-analysis

If few of the primary studies provide the estimates of reliability (and other relevant artifacts) required to correct for measurement error (and other artifacts), meta-analysis can nevertheless be carried out by use of such estimates from other credible sources—other relevant studies, test manuals, etc. This procedure is called artifact distribution meta-analysis (Hunter & Schmidt, 2004, chap.

4). Eq. (4) is the fundamental equation for this form of meta-analysis.

$$\hat{S}_\rho^2 = \left(\frac{\hat{\rho}}{\bar{r}}\right)^2 S_r^2 - \left(\frac{\hat{\rho}}{\bar{r}}\right)^2 E(S_{e_{r_i}}^2). \quad (4)$$

The term on the left side of Eq. (4) is the estimate of the variance of the population (parameter) disattenuated correlations. In this form of meta-analysis, it is not possible to directly compute the variance of the corrected correlations, because observed correlations are not corrected individually. However, the first term on the right side of Eq. (4) estimates this value as the product of the variance of the uncorrected correlations and the square of the mean correction factor. The correction factor is the factor by which the mean observed correlation,  $\bar{r}$  (the bare bones meta-analytic mean estimate), is increased when the mean true score population correlation,  $\hat{\rho}$ , is estimated, based on the correction for mean level of measurement error. This follows from the fact that the measurement error corrections increase the variance of the correlations by the square of this factor (Hunter & Schmidt, 2004, chap. 4). The basic statistical principle here is that if one multiplies any set of numbers by a constant, the standard deviation is multiplied by that constant and the variance is multiplied by the square of that constant. The second term on the right side of Eq. (4) estimates the sampling error variance in the same manner, based on the same principle. The corrections for measurement error increase sampling error variance, and the factor by which it is increased is again the square of the mean correction factor. In the second term on the right in Eq. (4), the sampling error variance is the weighted average of the sampling error variances in the individual (primary) studies included in the meta-analysis. The term on the left side of Eq. (4) ( $\hat{S}_\rho^2$ ) is the estimate of the variance of the disattenuated population correlations (i.e., the true score correlations). If this term is zero or very small, the final result is the (weighted) mean corrected correlation ( $\hat{\rho}$ ) and its standard deviation corrected for sampling error and measurement error ( $SD_\rho$ ). If the term on the left side of Eq. (4) ( $\hat{S}_\rho^2$ ) is large, there may be a search for moderators.

These are the basic principles and methods of first order psychometric meta-analysis. It is well accepted that meta-analysis has proven to be a critical and important methodological advance and has contributed greatly to the advancement of research progress and cumulative knowledge (e.g., see Bangert-Drowns, 1986; Chan & Arvey, 2012; DeGeest & Schmidt, 2010; McDaniel, 2007; Murphy & Newman, 2003; Sackett, 2003). However, it is nevertheless important to continue to strive for additional methodological advances in order to add additional useful tools for researchers that will help to further clarify the knowledge bases contained in our research literatures. In this connection, meta-analysis has been and will be a constantly evolving research integration tool (Schmidt, 2008). This is the motivation for the present paper on second order meta-analysis methods.

#### Second order meta-analysis: general principles

As mentioned briefly, second order meta-analysis has two important advantages over other approaches to interpreting multiple meta-analyses on the same question. First, second order meta-analysis allows estimation of the extent to which second order sampling error variance explains the differences between the mean effects across first order meta-analysis and allows estimation of the amount (if any) of true (i.e., non-artifactual) variance across these mean effect sizes. Second, second order meta-analysis allows one to compute the reliability of the differences between meta-analyses in mean effect sizes. This, in turn, allows more accurate estimation of the true mean effect sizes in each first order meta-analysis. No other method of interpreting multiple first order

meta-analyses allows either of these advantages. The nature and value of these advantages will become clearer as we present and examine the methods of second order meta-analysis.

These basic equations and principles of first order meta-analysis can be generalized to second order meta-analysis. We again use the correlation coefficient as the illustrative statistic; the equations for the  $d$  statistic are directly analogous. The input to a first order meta-analysis is an effect size estimate from each of the primary studies included in the meta-analysis. The input to a second order meta-analysis is, by contrast, the meta-analytic mean effect size estimate from each of the  $m$  meta-analyses included in the second order meta-analysis; that is,  $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3 \dots \hat{\rho}_m$ .

Among other purposes, second order meta-analysis provides a method for testing the reality of multiple hypothesized moderator variables simultaneously. A finding that second order sampling error accounts for all of the variability of the mean correlations or effect sizes across the individual meta-analyses suggests that the observed differences among the individual meta-analysis means do not represent real moderator effects. As discussed later, alternatives to second order meta-analysis cannot provide this information. In addition, as shown later, if second order sampling error accounts for less than 100 percent of the variance of the mean correlations or effect sizes across meta-analyses, the method of second order meta-analysis provides a useful means for increasing the accuracy of the estimates of the mean correlations in each individual meta-analysis.

It is important to note that second order meta-analysis requires the assumption that the different first order meta-analyses are statistically independent. Strictly speaking, the requirement of statistical independence means that the primary studies or samples contained in a first order meta-analysis should not also be included in any of the other first order meta-analyses. There are many situations in which this assumption is met. For instance, we later present an example application of second order meta-analysis in which the first order meta-analyses were all carried out in different countries and were therefore statistically independent (Oh, 2009). We present another example application in which multiple independent first order meta-analyses were carried out by the same research team (Mesmer-Magnus, DeChurch, Jimenez, Wildman, & Shuffler, 2011). Other examples of multiple independent meta-analyses can be found in the literature. The following are recent examples. Podsakoff, Bommer, Podsakoff, and MacKenzie (2006) examined industry type (manufacturing vs. service) as a moderator of the relationship between leader behavior and subordinate performance. Combs, Liu, Hall and Ketchen (2006) also conducted separate, statistically independent meta-analyses split out by type of industry. These subgroup meta-analyses had no studies in common and so were statistically independent. Van Iddekinge, Roth, Putka, and Lanivich (2011) conducted separate statistically independent meta-analyses on job applicants and job incumbents and also on predictive and concurrent validity studies in examining the relation between vocational interests and job performance. Wang, Oh, Courtright, and Colbert (2011) conducted separate, statistically independent meta-analyses on public and private organizations as well as on North American and East Asian samples.

However, in many cases this assumption is not met or is only partially met. Cooper and Koenka (2012) suggest that simply minimizing the lack of independence might be the best that can sometimes be expected (p. 458). Tracz, Elmore, and Pohlmann (1992), in an extensive simulation study, found that violations of independence had little or no effect on meta-analytic results. We discuss this question in more detail later in the paper. Finally, in many cases the equations presented below can be further simplified; however, we present them in pre-simplified form to facilitate understanding of the logic reflected in the equations.

### Second order meta-analysis of bare bones meta-analyses

Eq. (5) is the fundamental equation when the first order meta-analyses entering the second order meta-analysis have corrected only for sampling error:

$$\hat{\sigma}_{\hat{\rho}_{xy}}^2 = S_r^2 - E(S_{e_{r_i}}^2), \tag{5}$$

where the term on the left side of the equation is the estimate of the population variance of the uncorrected mean correlations ( $\hat{\rho}_{xy}$ ) across the meta-analyses after second order sampling error has been subtracted out. The first term on the right side of Eq. (5) is the weighted variance of the mean correlations across the  $m$  meta-analyses, computed as follows:

$$S_r^2 = \frac{\sum_1^m w_i (\hat{r}_i - \hat{\bar{r}})^2}{\sum_1^m w_i}, \tag{5a}$$

where

$$\hat{\bar{r}} = \frac{\sum_1^m w_i \hat{r}_i}{\sum_1^m w_i}; \tag{5b}$$

and

$$w_i = \left( \frac{S_{r_i}^2}{k_i} \right)^{-1}, \tag{5c}$$

and where  $S_{r_i}^2$  is the variance of the observed correlations ( $r$ s) in the  $i$ th meta-analysis,  $\hat{r}_i$  is the estimate of the mean effect size for the  $i$ th meta-analysis,  $\hat{\bar{r}}$  is the estimate of the (weighted) grand mean effect size across the  $m$  meta-analyses,  $k_i$  is the number of primary studies included in the  $i$ th meta-analysis, and the  $w_i$  is the weight applied to the  $i$ th meta-analysis. The second term on the right side of Eq. (5) is the expected (weighted average) second order sampling error variance across the  $m$  meta-analyses:

$$E(S_{e_{r_i}}^2) = \frac{\sum_1^m \left( w_i \frac{S_{r_i}^2}{k_i} \right)}{\sum_1^m w_i}. \tag{5d}$$

Eq. (5d) reduces to Eq. (5e):

$$E(S_{e_{r_i}}^2) = m \frac{1}{\sum_1^m w_i}. \tag{5e}$$

To summarize, each meta-analysis will have reported a mean uncorrected (i.e., mean observed) correlation,  $\hat{r}_i$ . The first term on the right in Eq. (5) is the weighted variance of these mean correlations. This computation is shown in Eqs. (5a) and (5b). The weights ( $w_i$ ) used in Eqs. 5a, 5b, 5d, and 5e are as defined in Eq. (5c). Each weight is the inverse of the random effect (RE) sampling error variance for the mean correlation in the  $i$ th meta-analysis (Schmidt et al., 2009). The second term on the right in Eq. (5) is the sampling error variance of these mean correlations. Each of the meta-analyses will have reported the variance of the observed correlations in that meta-analysis. Dividing each such variance by  $k_i$  (the number of studies in that meta-analysis) yields the RE sampling error variance of the mean  $r$  ( $\hat{r}_i$ ) in that meta-analysis (Schmidt et al., 2009). [This reflects the well known principle that the sampling error variance of the mean of any set of scores is the variance of the scores divided by the number of scores (and the standard error of the mean is the square root of this value).] The weighted average of these values across the  $m$  meta-analyses estimates the RE sampling error variance of the mean  $r$ s as a group, as shown in Eq. (5d) and Eq. (5e). The square root of this value divided by the square root of  $m$  is the standard error ( $SE_{\hat{r}}$ ) and can be used to put confidence intervals around the estimate of the (weighted) grand mean ( $\hat{\bar{r}}$ ; computed in Eq. (5b)). Also, using the square root of the value on

the left side of Eq. (5) ( $\hat{\sigma}_{\rho_{xy}}$ ) one can construct a credibility interval around the grand mean correlation across the  $m$  meta-analyses, within which a given percentage of the first order population meta-analytic (mean) effect sizes ( $\hat{\rho}_{xy}$ ) is expected to lie (Hunter & Schmidt, 2004, chapter 3). For example, 80% would be expected to lie within the 80% credibility interval. If the value on the left side of Eq. (5) is zero, the conclusion is that the mean population correlation values are the same across the meta-analyses. In that case, all the observed variance is accounted for by second order sampling error, and the conclusion is that there are no moderators. If it is greater than zero, one can compute the proportion of variance between-meta-analyses that is due to second order sampling error. This is computed as the ratio of the second term on the right side of Eq. (5) to the first term on the right side, i.e.:

$$\text{ProportionVar} = E(S_{e_i}^2) / S_{\hat{\rho}}^2, \tag{5f}$$

and  $1 - \text{ProportionVar}$  denotes the proportion of the variance across first order meta-analytic (bare bones) mean correlations that is “true” variance (i.e., variance not due to second order sampling error). As such, this number is the reliability of the meta-analytic correlations (considered as a set of values, one for each first order meta-analysis) (Hunter & Schmidt, 2004). This follows because reliability is the proportion of total variance that is true variance (Ghiselli, 1964; Magnusson, 1966; Nunnally & Bernstein, 1994). As discussed later, this value can be used to produce enhanced accuracy for estimates of these mean (meta-analytic) correlations from the first order meta-analyses by regressing them towards the value of the grand mean correlation (the mean across the first order meta-analyses). Both of these analyses are unique to second order meta-analysis and cannot be performed using other analysis methods.

*Second order meta-analysis when correlations have been individually corrected*

Measurement error is present in all research and it biases all relationships examined in research. Therefore it is important to include corrections for these biases. One approach in meta-analysis is to correct each correlation individually for the downward bias created by measurement error (Hunter & Schmidt, 2004, chap. 3). When the first order meta-analyses entering the second order meta-analysis have corrected each correlation individually for measurement error (and range restriction and dichotomization, if applicable), the fundamental equation for second order meta-analysis is:

$$\hat{\sigma}_{\rho}^2 = S_{\hat{\rho}}^2 - E(S_{e_i}^2), \tag{6}$$

where the term on the left in Eq. (6) is the estimate of the actual (non-artifactual) variance across the  $m$  meta-analyses of the population mean disattenuated correlations ( $\hat{\rho}$ ); that is, the variance after variance due to second order sampling error has been subtracted out. The first term on the right side of Eq. (6) is the variance of the mean individually corrected correlations across the  $m$  meta-analyses, computed as follows:

$$S_{\hat{\rho}}^2 = \sum_1^m w_i^* (\hat{\rho}_i - \hat{\rho})^2 / \sum_1^m w_i^*; \tag{6a}$$

where

$$\hat{\rho} = \sum_1^m w_i^* \hat{\rho}_i / \sum_1^m w_i^*; \tag{6b}$$

and

$$w_i^* = \left( \frac{S_{r_{c_i}}^2}{k_i} \right)^{-1}, \tag{6c}$$

and where  $S_{r_{c_i}}^2$  is the weighted variance of the disattenuated (individually corrected) correlations in the  $i$ th meta-analysis,  $\hat{\rho}_i$  is the mean meta-analytic disattenuated correlation in that meta-analysis,  $\hat{\rho}$  is the (weighted) grand mean effect size across the  $m$  meta-analyses,  $k_i$  is the number of primary studies included in the  $i$ th meta-analysis, and the  $w_i^*$  is the weight applied to the  $i$ th meta-analysis. The second term on the right side of Eq. (6) is the weighted average second order sampling error variance across the  $m$  meta-analyses:

$$E(S_{e_i}^2) = \sum_1^m w_i^* \left( \frac{S_{r_{c_i}}^2}{k_i} \right) / \sum_1^m w_i^*. \tag{6d}$$

Eq. (6d) reduces to Eq. (6e):

$$E(S_{e_i}^2) = m / \sum_1^m w_i^* \tag{6e}$$

where the  $w_i^*$  are as defined in Eq. (6c).

To summarize, each first order meta-analysis will have reported an estimate of the mean disattenuated correlation (the meta-analytic mean correlation,  $\hat{\rho}_i$ ). The first term on the right side of Eq. (6) is the variance of these meta-analytic mean correlations across first order meta-analyses. This computation is shown in Eqs. (6a) and (6b). Eq. (6c) shows the weights that are used in Eqs. (6a) and (6b). The second term on the right side of Eq. (6) is the expected value of the second order sampling error variance of these meta-analytic correlations. Each meta-analysis will have reported an estimate of the variance of the corrected correlations it included, preferably to four decimal places, for precision.<sup>3</sup> Dividing this value by  $k$  (the number of studies in the meta-analysis), yields the RE sampling error variance of the meta-analytic correlation for that meta-analysis (Schmidt et al., 2009). [As noted earlier, this reflects the well known statistical principle that the sampling error variance of the mean of any set of scores is the variance of the scores divided by the number of scores (and the standard error of the mean is the square root of this value).] As shown in Eqs. (6d) and (6e), the weighted mean of these values across the  $m$  meta-analyses yields the second order sampling error variance needed in Eq. (6). The square root of this value divided by the square root of  $m$  is the standard error ( $SE_{\hat{\rho}}$ ) and can be used to put confidence intervals around the grand mean ( $\hat{\rho}$ ; shown in Eq. (6b)).

The term on the left side of Eq. (6) is the estimate of the actual (non-artifactual) variance across meta-analysis of the population mean disattenuated correlations (the  $\hat{\rho}_i$ ); that is, the variance across first order meta-analytic estimates after removal of variance due to second order sampling error. Using the square root of this value ( $\hat{\sigma}_{\rho}$ ), credibility intervals can be placed around the grand mean computed in Eq. (6b). For example, 80% of population mean values are expected to lie within in the 80% credibility interval.

If the value on the left side of Eq. (6) is zero, the indicated conclusion is that the mean population correlation values are the same across the multiple meta-analyses. All the variance is accounted for by second order sampling error. If this value is greater than zero, one can compute the proportion of the between-meta-analyses variance that is explained by second order sampling error. This is computed as ratio of the second term on the right side of Eq. (6)

<sup>3</sup> If the variance (or standard deviation) of observed effect sizes is reported only up to two decimal places, second order meta-analysis results may be less accurate. Therefore we urge meta-analysts report the variance of the corrected correlations (when individual correction methods are used) or the variance of the observed correlations (when artifact distribution methods are used) to more than two decimal places or make it publicly available (e.g., digital object identifier) wherever possible so that later second order meta-analyses can be conducted more precisely. Hurtz and Donovan (2000), Salgado (1997) and Salgado (1998) are some excellent examples of this sort. More importantly, we urge the meta-analysts to report the variance (or standard deviation) of observed effect sizes to begin with. Unfortunately, we have often found that many meta-analyses published in top tier journals do not report the variance (or standard deviation) of observed effect sizes (see Carlson & Ji, 2011 for how to effectively report and cite first order meta-analytic results).

to the first term on the right side, i.e.:

$$\text{ProportionVar} = E(S_{e_{\hat{\rho}_i}}^2) / S_{\hat{\rho}}^2, \tag{6f}$$

and  $1 - \text{ProportionVar}$  denotes the proportion of the variance across the first order meta-analysis mean population correlation values that is true variance (i.e., variance not due to second order sampling error). As such, this number is the reliability of the estimated mean first order population correlations (Hunter & Schmidt, 2004), because reliability is the proportion of total variance that is true variance (Ghiselli, 1964; Nunnally & Bernstein, 1994). As illustrated later, this value can be used to refine the estimates of these first order meta-analysis mean values by regressing them towards the value of the grand mean disattenuated correlation (the mean across the  $m$  meta-analyses, computed in Eq. (6b)). This procedure is illustrated in the example applications presented in the next section of this paper. Both these analyses are unique to second order meta-analysis and are not available from other methods. In addition, when  $S_{\hat{\rho}}^2$  is zero, the ProportionVar is 100% and the reliability of the vector of  $m$  first order meta-analytic mean estimates is zero (e.g., Conscientiousness, in Column 11 of Table 2). This is the same as the situation in which all examinees get the same score on a test, making the reliability of the test zero.

We note that ProportionVar is less informative when the observed variance of the meta-analytic mean correlations across the  $m$  meta-analyses ( $S_{\hat{\rho}}^2$ ) is close to zero. A percent-based estimate can be misleading when it is interpreted blindly without considering the size of its denominator. For example, a ProportionVar figure of 50% could be .1000/.2000 or .00010/.00020. The latter case would not suggest the existence of moderator(s) across  $m$  first order meta-analytic mean estimates, given the tiny amount of observed variation to begin with and the even smaller amount of non-artifactual variance. For purposes of detecting the likely presence of moderators across the  $m$  first order meta-analytic mean estimates, the absolute amount of true variance across  $m$  first order meta-analytic mean estimates ( $\hat{\sigma}_{\hat{\rho}}^2$ ) (or even better, its square root, the SD) can be more important than the relative percent of variance attributable to second order sampling error. (This same principle applies within first-order meta-analyses.) We suggest that meta-analysts consider both estimates.

*Second order meta-analysis with artifact distribution meta-analyses*

Often the information needed to correct each correlation individually for measurement error is unavailable for many or most of the studies. In such literatures, meta-analysis can nevertheless correct for measurement error by use of measurement error estimates (reliability estimates) from other credible sources, as indicated earlier. This method of meta-analysis is called artifact distribution meta-analysis (Hunter & Schmidt, 2004, chap. 4). Eq. (7) is the fundamental equation for second order meta-analysis when the first order meta-analyses have applied the artifact distribution method of meta-analysis.

$$\hat{\sigma}_{\hat{\rho}}^2 = S_{\hat{\rho}}^2 - E(S_{e_{\hat{\rho}_i}}^2), \tag{7}$$

where the term on the left side of Eq. (7) is the estimate of the non-artifactual variance of the population meta-analytic (disattenuated) correlations (population parameter values) across the  $m$  first order meta-analyses. This is the variance remaining after variance due to second order sampling error has been subtracted out. The first term on the right side of Eq. (7) is the variance of the mean disattenuated correlations across the  $m$  meta-analyses, computed as follows:

$$S_{\hat{\rho}}^2 = \sum_1^m w_i^{**} (\hat{\rho}_i - \hat{\hat{\rho}})^2 / \sum_1^m w_i^{**}; \tag{7a}$$

where

$$\hat{\hat{\rho}} = \sum_1^m w_i^{**} \hat{\rho}_i / \sum_1^m w_i^{**}; \tag{7b}$$

and

$$w_i^{**} = \left[ \left( \frac{\hat{\rho}_i}{\bar{r}_i} \right)^2 \left( \frac{S_{r_i}^2}{k_i} \right) \right]^{-1}, \tag{7c}$$

and where  $S_{r_i}^2$  is the variance of the observed correlations within a given meta-analysis,  $\hat{\rho}_i$  is the mean disattenuated correlation in that meta-analysis,  $\bar{r}_i$  is the meta-analytic (bare bones) mean correlation in that meta-analysis,  $\hat{\hat{\rho}}$  is the (weighted) grand mean effect size across the  $m$  meta-analyses,  $k_i$  is the number of primary studies included in the  $i$ th meta-analysis, and  $w_i^{**}$  is the weight applied to the  $i$ th meta-analysis. The second term on the right side of Eq. (7) is the weighted average second order sampling error variance across the  $m$  meta-analyses:

$$E(S_{e_{\hat{\rho}_i}}^2) = \sum_1^m \left[ w_i^{**} \left( \frac{\hat{\rho}_i}{\bar{r}_i} \right)^2 \frac{S_{r_i}^2}{k_i} \right] / \sum_1^m w_i^{**}. \tag{7d}$$

Eq. (7d) reduces to Eq. (7e):

$$E(S_{e_{\hat{\rho}_i}}^2) = m / \sum_1^m w_i^{**}. \tag{7e}$$

The  $w_i^{**}$  are as defined in Eq. (7c). Eq. (7) has the same form as Eq. (6) but some of the terms in it are estimated differently, so some explanation is indicated. The first term on the right side of Eq. (7) is the computed variance across the meta-analyses of the first order meta-analytic mean disattenuated population correlations. Computation of this value is shown in Eqs. (7a) and (7b). Eq. (7c) shows the weights that are applied in Eqs. (7a) and (7b). The second term on the right in Eq. (7) is the sampling error variance of these estimates. As shown in Eqs. (7d) and (7e), this sampling error is estimated as the weighted average across meta-analyses of the product of the square of the mean correction factor and the mean sampling error variance of the bare bones (uncorrected) meta-analytic correlations ( $S_{e_r}^2$ ; see Eq. (5d)). Each meta-analysis will have reported the variance of the observed correlations it included. Dividing this variance by  $k$  (the number of studies in the meta-analysis) yields the RE sampling error variance of the mean of the observed (uncorrected) correlations in that meta-analysis. As shown in Eqs. (7d) and (7e), the weighted average of the product of these values and the square of the correction factors across the  $m$  meta-analyses is the random effects sampling error variance estimate needed for Eq. (7) (Hunter & Schmidt, 2004, chap. 4). This is based on the well known principle that if one multiplies a distribution of scores by a constant, the standard deviation is multiplied by that constant and the variance is multiplied by the square of that constant. Here the constant is the mean measurement error correction ( $\hat{\rho}_i / \bar{r}_i$ ). The square root of the value of the left side of Eq. (7d) divided by the square root of  $m$  is the standard error ( $SE_{\hat{\rho}}$ ) and can be used to put confidence intervals around the grand mean ( $\hat{\hat{\rho}}$ ; computed in Eq. (7b)).

The value on the left side of Eq. (7) is the estimate of the non-artifactual variance of the population disattenuated correlations across the  $m$  meta-analyses. This is the variance remaining after subtraction of variance due to second order sampling error. When this value is negative (i.e., second order sampling error variance is greater than the observed variance across the first order meta-ana-

**Table 1**  
Summary of three approaches to second order meta-analysis.

	Bare bones second order meta-analysis	Psychometric second order meta-analysis	Artifact distribution first order meta-analysis
Input	Uncorrected first order meta-analysis estimates ( $\hat{\rho}_{xy}$ )	Individually corrected first order meta-analysis estimates ( $\hat{\rho}$ )	Artifact distribution first order meta-analysis estimates ( $\hat{\rho}$ )
General principles	$\hat{\sigma}_{\rho_{xy}}^2 = S_r^2 - E(S_{e_{r_i}}^2)$ $\hat{\sigma}_{\rho_{xy}}^2$ is the estimate of the population variance of the uncorrected mean correlations ( $\hat{\rho}_{xy}$ )	$\hat{\sigma}_{\rho}^2 = S_{\rho}^2 - E(S_{e_{\rho_i}}^2)$ $\hat{\sigma}_{\rho}^2$ is the estimate of the actual (non-artifactual) variance across the $m$ meta-analyses of the population mean (individually) disattenuated correlations ( $\hat{\rho}$ )	$\hat{\sigma}_{\rho}^2 = S_{\rho}^2 - E(S_{e_{\rho_i}}^2)$ $\hat{\sigma}_{\rho}^2$ is the estimate of the actual (non-artifactual) variance across the $m$ meta-analyses of the population mean disattenuated correlations ( $\hat{\rho}$ )
Second order sampling error	$E(S_{e_{r_i}}^2) = \sum_1^m (w_i \frac{S_{r_i}^2}{k_i}) / \sum_1^m w_i$	$E(S_{e_{\rho_i}}^2) = \sum_1^m w_i^* (\frac{S_{r_i}^2}{k_i}) / \sum_1^m w_i^*$	$E(S_{e_{\rho_i}}^2) = \sum_1^m [w_i^* (\frac{\hat{\rho}_i}{r_i})^2 \frac{S_{r_i}^2}{k_i}] / \sum_1^m w_i^*$
Observed between-first-order meta-analyses variance	$S_r^2 = \sum_1^m w_i (\hat{r}_i - \hat{r})^2 / \sum_1^m w_i$	$S_{\rho}^2 = \sum_1^m w_i^* (\hat{\rho}_i - \hat{\rho})^2 / \sum_1^m w_i^*$	$S_{\rho}^2 = \sum_1^m w_i^* (\hat{\rho}_i - \hat{\rho})^2 / \sum_1^m w_i^*$
Second order grand mean	$\hat{r} = \sum_1^m w_i \hat{r}_i / \sum_1^m w_i$	$\hat{\rho} = \sum_1^m w_i^* \hat{\rho}_i / \sum_1^m w_i^*$	$\hat{\rho} = \sum_1^m w_i^* \hat{\rho}_i / \sum_1^m w_i^*$
Weight for each first order meta-analysis	$w_i = (\frac{S_{r_i}^2}{k_i})^{-1}$ $S_{r_i}^2$ is the variance of the observed correlations ( $r$ ) in the $i$ th meta-analysis, $\hat{r}_i$ is the estimate of the mean effect size in that meta-analysis, $k_i$ is the number of primary studies included in the $i$ th meta-analysis, and the $w$ is the weight applied to the $i$ th meta-analysis	$w_i^* = (\frac{S_{r_i}^2}{k_i})^{-1}$ $S_{r_i}^2$ is the weighted variance of the disattenuated (corrected) correlations ( $r_c$ ) in the $i$ th meta-analysis, $\hat{\rho}_i$ is the mean disattenuated correlation in that meta-analysis, $k_i$ is the number of primary studies included in the $i$ th meta-analysis, and the $w_i^*$ is the weight applied to $i$ th meta-analysis	$w_i^* = [(\frac{\hat{\rho}_i}{r_i})^2 (\frac{S_{r_i}^2}{k_i})]^{-1}$ $S_{r_i}^2$ is the variance of the observed correlations ( $r$ ) within a given meta-analysis, $\hat{\rho}_i$ is the mean disattenuated correlation in that meta-analysis, $r_i$ is the meta-analytic (bare bones) correlation in that meta-analysis, $k_i$ is the number of primary studies included in the $i$ th meta-analysis, the $w_i^*$ is the weight applied to the $i$ th meta-analysis
ProportionVar	$E(S_{e_{r_i}}^2) / S_r^2$ This is the proportion of the variance across the first order bare bones meta-analytic means that is due to second order sampling error	$E(S_{e_{\rho_i}}^2) / S_{\rho}^2$ This is the proportion of the variance across the first order psychometric meta-analytic means that is due to second order sampling error	$E(S_{e_{\rho_i}}^2) / S_{\rho}^2$ This is the proportion of the variance across the first order psychometric meta-analytic means that is due to second order sampling error

Note. 1 – ProportionVar = Reliability of the first order meta-analytic correlations.

lytic mean estimates), it is set to zero. Using the square root of this value ( $\hat{\sigma}_{\rho}$ ), credibility intervals around the grand mean correlation can be computed, as described earlier. If the value on the left side of Eq. (7) is zero, the indicated conclusion is that these mean population correlations are the same across the  $m$  meta-analyses. All variance is accounted for by second order sampling error, leading to the conclusion that there are no moderators. If this value is greater than zero, one can compute the proportion of between-meta-analysis variance that is accounted for by second order sampling error variance. This is computed as the ratio of the second term on the right side of Eq. (7) to the first term, i.e.:

$$\text{ProportionVar} = E(S_{e_{\rho_i}}^2) / S_{\rho}^2, \tag{7f}$$

where 1 – ProportionVar denotes the proportion of the variance of the population disattenuated correlations that is true variance (i.e., variance not due to second order sampling error). Because of this, this number is the reliability of the vector of mean corrected correlations across the  $m$  first order meta-analyses. This reliability reflects the extent to which the mean first order corrected correlations discriminate between the first order meta-analysis results. This reliability value can be used to increase the accuracy of the estimates of first order meta-analytic (disattenuated) correlations from the individual first order meta-analyses by regressing these estimates towards the grand correlation mean (the second order mean across the  $m$  meta-analyses; shown in Eq. (7b)), as illustrated later in the example applications.

We are aware that the procedures and concepts of second order meta-analysis are complex. For this reason, Table 1 presents a convenient summary of the equations for all three approaches to sec-

ond order meta-analysis. The entries in Table 1 are annotated for increased clarity.

*Mixed second order meta-analysis*

In some cases, some of the first order meta-analyses might have corrected each correlation individually while others applied the artifact distribution method. How, then, should the second order meta-analysis be conducted? The meta-analyses that corrected each coefficient individually can be “converted” to artifact distribution meta-analyses and the Equations for second order artifact distribution meta-analysis can be applied to all the first order meta-analysis. The quantities needed in these equations (Eqs. (7) and (7a)–(7f)) are typically reported in meta-analyses that have corrected each correlation individually, making this conversion possible.

**Three illustrative applications**

*Case 1: Cross-cultural validity generalization analysis*

Oh (2009) examined the criterion-related validity of self-report personality measures explicitly developed to assess the Big Five traits for predicting supervisor ratings of job performance in four East Asian countries including Korea, China, Taiwan, and Singapore.<sup>4</sup> Oh (2009) expected different results across these countries

<sup>4</sup> Japan was also initially included but Oh (2009) found that in Japan, unlike the other East Asian countries, the personality measures used were not designed to measure Big Five personality traits. Hence Japan was excluded to avoid confounding due to the differences across countries in personality measures (Spector, 2001, p. 23).

**Table 2**  
Second order meta-analysis of operational validities of big five personality measures across East Asian countries (Oh, 2009).

Predictor	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Country	<i>k</i>	$\bar{r}_i$	$S_r^2$	$\hat{\rho}_i$	$S_{e_{\rho_i}}^2$	$\hat{\rho}$	$E(S_{e_{\rho_i}}^2)$	$S_{\rho}^2$	$\sigma_{\rho}^2$	ProVar	$r_{\rho\rho}$	$\hat{\rho}_{tr}$
<i>Conscientiousness</i>												
Korea	14	.13	.00501	.19	.00072	.21	.00124	.00107	.00000	1.00	.00	.21
Taiwan	3	.26	.01331	.36	.00853		(.035)	(.033)	(.000)			.21
China	3	.17	.02325	.23	.01490							.21
Singapore	4	.15	.00126	.21	.00061							.21
<i>Emotional Stability</i>												
Korea	14	.08	.00677	.11	.00098	.17	.00221	.00647	.00427	.34	.66	.13
Taiwan	2	-.03	.01584	-.04	.01524		(.047)	(.080)	(.065)			.03
China	3	.17	.01074	.23	.00688							.21
Singapore	4	.19	.00355	.26	.00171							.23
<i>Agreeableness</i>												
Korea	13	.04	.00468	.05	.00066	.10	.00144	.00500	.00357	.29	.71	.07
Taiwan	3	.21	.00706	.30	.00453		(.038)	(.071)	(.060)			.24
China	3	.14	.01708	.20	.01095							.17
Singapore	3	.10	.00163	.13	.00104							.12
<i>Extraversion</i>												
Korea	14	.06	.00491	.09	.00074	.17	.00173	.01066	.00893	.16	.84	.10
Taiwan	3	.14	.01160	.19	.00743		(.042)	(.103)	(.095)			.19
China	3	.21	.03720	.30	.02385							.28
Singapore	4	.22	.00262	.31	.00126							.28
<i>Openness</i>												
Korea	14	.01	.00881	.02	.00119	.10	.00229	.01087	.00858	.21	.79	.04
Taiwan	2	.24	.00460	.34	.00442		(.048)	(.104)	(.093)			.29
China	3	.15	.08494	.21	.05445							.19
Singapore	3	.10	.00235	.13	.00151							.13

Note. Columns (1) through (4) are input values (italicized) available from first order meta-analyses. (1) Number of samples; (2) Sample size weighted mean observed validity; (3) Sample size weighted observed variance across observed validities; (4) First order meta-analytic mean validity estimates; (5) Second order sampling error variance for each first order meta-analytic validity estimate (see discussion of Eq. (7d)); (6) Second order, grand mean validity estimates (Eq. (7b)); (7) Expected (average) second order sampling error variance (Eq. (7d)) and standard error (in parentheses); (8) Observed variance and SD (in parentheses) across first order mean operational validity estimates (Eqs. (7a), (7b), and (7c)); (9) Estimated true variance and SD (in parentheses) across first order mean operational validity estimates after expected second order sampling error variance is subtracted out from the observed variance (Eq. (7)); negative values are set to zero; (10) The proportion (percentage if multiplied by 100) of the observed variance across first order mean operational validity estimates that is due to second order sampling error variance; values greater than 1 are set to 1; (11) The reliability of the first order meta-analytic validity vectors; these values are computed as 1 minus the values in Column 10; (12) Regressed first order validity estimates based on the reliability of the original validity vectors shown in Column 11.

because cultural values, industrialization levels, and management practices of these countries are different to some extent (Hofstede, 2010). He conducted a separate meta-analysis for each country for each of the Big Five personality measures, with each meta-analysis being based on only the primary studies that had been conducted in that country. Hence the meta-analytic results across countries are statistically independent of each other, meeting this assumption of second order meta-analysis.

The major first order meta-analytic results from Oh (2009) are presented in Columns 1 through 4 of Table 2; they are the input to the subsequent second order meta-analyses, shown in columns 5 through 12 of Table 2. All the first order meta-analyses were based on the artifact distribution method. Visual inspection of the input data (first order meta-analytic results) in Table 2 suggested that there were indeed substantial differences in validity across countries. For example, the personality trait of Emotional Stability appears to be most valid in Singapore (.26) and least valid in Taiwan (-.04). So, initial examination of the results suggested a country by validity interaction (a moderating effect of country/culture on validity). However, each meta-analytic estimate (correlation) contains some residual sampling error given that sample sizes are less than infinite (*k* is often less than 10); that is, there is second order sampling error in these estimates. The values for this second order sampling error are presented in Column 5 of Table 2 next to each mean first order mean operational validity estimate in Column 4. For each first order meta-analysis, these values are computed as the inverse of Eq. (7c).

To determine the degree to which the differences in operational validity across countries shown in Table 2 were due to second or-

der sampling error, we conducted a second order meta-analysis for each Big Five personality measure across countries, using the methods presented earlier in this paper for meta-analyses based on the artifact distribution method. Note that this second order meta-analysis was conducted “using only the first order meta-analytic results” reported in Oh (2009). The results are shown in columns 5 through 12 of Table 2. Column 6 shows the weighted grand mean operational validity across the countries for a given Big Five measure (from Eq. (7b)) and in Column 7, the expected second order sampling error associated with each grand mean (from Eq. (7d)) (along with its SD [in parenthesis]). In Column 8, the observed variance and SD (in parentheses) of the mean operational validity estimates across countries are shown; these values are computed using Eqs. 7a, 7b, and 7c. Column 9 shows these values after second order sampling error is subtracted out. These values are computed using Eq. (7), and they estimate the actual/true (i.e., non-artifactual) variability of the operational validities across countries. As expected, these values are reduced in comparison to the observed values of variance in the row above them. Column 10 presents the proportion of the variance that is due to second order sampling error. This value is computed using Eq. (7f). If multiplied by 100, this represents the percentage of between-country (or between-meta-analyses) variance that is due to second order sampling error. Column 11 presents the reliabilities of the validity vectors shown in Column 4 of Table 2. These values are computed as 1.00 minus the values in Column 10 and represent the proportion of the variance that is not due to second order sampling error (i.e., they represent the proportion of between-country variance that is true variance).



On average across the Big Five personality traits, 40% of the cross-country variance is explained by second order sampling error. The results for the personality trait of Conscientiousness are of particular interest, because across multiple meta-analyses conducted in the U.S. and Europe, Conscientiousness has been found to be the most valid of the Big Five personality traits for the prediction of job performance (e.g., see Barrick et al., 2001; Schmidt, Shaffer, & Oh, 2008). The second order meta-analysis results show that all the variability of the operational validities for Conscientiousness across these East Asian countries is due to second order sampling error. This suggests that the grand mean operational validity across these countries (.21) is the best estimate for each country and other East Asian countries not included.

For the other four Big Five personality traits, removal of second order sampling error reduced variability but not to zero. For Emotional Stability measures, second order sampling error accounted for 34% of the observed between-country variance, indicating a correlation of .58 between country-specific mean operational validity estimates and second order sampling errors ( $\sqrt{.34} = .58$ ). For Extraversion measures, 16% of the variance is accounted for, indicating a correlation of .40 between between-country mean operational validity estimates and second order sampling errors. For Agreeableness measures, these values are 29% ( $r = .54$ ) and for Openness measures, these values are 21% ( $r = .46$ ). As implied above, for Conscientiousness measures, this correlation is 1.00, because all the variance is due to second order sampling error.

As noted earlier, the reliability values (shown in column 11 of Table 2 for our example) can be used to increase the accuracy of estimates of the first order meta-analytic mean correlation values obtained in the  $m$  individual first order meta-analyses. We explain and illustrate this process using the data in our example applications of second order meta-analysis. This process is directly analogous to the estimation of an individual's true score on a measure from his/her observed score by using the reliability coefficient to regress the obtained score towards the mean of the group. The equation is (Ghiselli, 1964; Magnusson, 1966; Nunnally & Bernstein, 1994):

$$\hat{T}_i = r_{xx}(X_i - \bar{X}) + \bar{X},$$

where  $\hat{T}_i$  is the estimate of true score on the measure for Person  $i$ ,  $r_{xx}$  is the reliability (the ratio of true score variance to observed score variance for the measure),  $X_i$  is the person's observed score, and  $\bar{X}$  is the mean of the observed scores across persons. It can be shown that the above equation is the linear regression equation (Magnusson, 1966) for predicting true score from observed score and application of the regression equation improves accuracy. When used in psychometrics, the error variance in question is measurement error variance. In second order meta-analysis, second order sampling error variance functions in the same way as measurement error. Just as a measurement error is the random deviation of a person's observed score from his/her true score, so a second order sampling error is the random deviation of the computed first order mean meta-analytic correlation in a meta-analysis from its actual population parameter value. Thus second order sampling error functions analogously to measurement error here. Hence, the above equation can be translated to Eq. (7g) below:

$$\hat{\rho}_{ir} = r_{\rho\rho}(\hat{\rho}_i - \hat{\rho}) + \hat{\rho}, \quad (7g)$$

where  $\hat{\rho}_{ir}$  is the regressed  $i$ th first order meta-analytic estimate adjusted for unreliability in the first order meta-analytic values,  $r_{\rho\rho}$  is the reliability of  $i$  first order meta-analytic values (vectors),  $\hat{\rho}$  is the second order, grand mean meta-analytic estimate, and  $\hat{\rho}_i$  is the  $i$ th first order meta-analytic estimate.

Columns 4 and 12 of Table 2, respectively, show the original first order meta-analytic mean operational validities as reported in the  $m$  meta-analyses (column 4) and the more accurate values

produced by application of Eq. (7g) (regressed values; column 12). For Conscientiousness, the reliability value is .00, and so the estimate for each country is equal to the grand mean value of .21. For the other Big Five measures, the regressed values are less variable across countries than the original reported values. This reduction in variability is greatest for Emotional Stability. It is interesting to note that the negative value for Emotional Stability in the Taiwan meta-analysis (−.04) yields a regressed value that is positive (.03), indicating that the initial negative value was due to second order sampling error.

It can be seen in this example application that the second order meta-analysis indicates that the deep structure underlying the data is simpler than the surface structure particularly for Conscientiousness. In the case of Conscientiousness, the second order meta-analyses suggest a much more parsimonious explanation than did examination of the first order meta-analysis results. A single value (.21) for the validity of Conscientiousness across the four East Asian countries is simpler and more parsimonious than different values for different countries/cultures. That is, the use of second order meta-analysis as a tool for cross-cultural/national validity generalization reveals that the principle of parsimony (Occam's razor) applies to the validity of Conscientiousness. Without the use of second order meta-analysis, we are not able to detect "the lies data tell" (Schmidt, 2010), resulting in the erroneous acceptance of the less parsimonious conclusion that the operational validity of Conscientiousness differs across the East Asian countries, ranging from .19 to .36.

In sum, first order meta-analysis is a useful tool to test hypotheses about within-culture/nation variability, and second order meta-analysis is a useful tool to test hypotheses about cross-culture/nation variability. As discussed and shown later, second order meta-analysis can directly test certain hypotheses about cross-nation variability that first order meta-analysis cannot test.

#### Case 2: differential validity analysis

Dudley, Orvis, Lebiecki, and Cortina (2006) examined the criterion-related validity (in the form of true score correlations) of four major facet measures of Conscientious (achievement, dependability, order, and cautiousness) for predicting supervisor ratings of various job performance criteria (global performance, task performance, job dedication, interpersonal facilitation, and counterproductive work behavior). All meta-analytic estimates were computed using the artifact distribution psychometric correction procedures. However, meta-analytic results across job performance criteria are not completely statistically independent though mostly so; there were small numbers of overlapping studies across criteria (in particular, between interpersonal facilitation and job dedication). Hence, first order meta-analytic results for a given criterion mostly, though not completely, met the assumptions of second order meta-analysis (see also the section below). In fact, this is a typical case for most if not all second order meta-analyses (Cooper & Koenka, 2012). It is well known that minor violations of statistical assumptions do not change research conclusions (Cooper & Koenka, 2012). Tracz et al. (1992) used simulations to examine the effect of violations of independence on the results of first order meta-analyses. Surprisingly, they found that that means, standard deviations, and confidence intervals were almost identical under conditions of independence and substantial violations of independence. We could locate no other studies examining the effects of violations of independence on meta-analytic results.

To determine the degree to which the differences in true score correlation across job performance criteria for a given Conscientiousness facet were due to second order sampling error, we conducted a second order meta-analysis for each Conscientiousness facet. Note that this second order meta-analysis was conducted using only the first order meta-analytic results reported in Dudley,

**Table 3**  
Second order meta-analysis of true score correlations of facet measures of conscientiousness with performance across performance criteria (Dudley et al., 2006).

Predictor Criterion	(1) <i>k</i>	(2) $\bar{r}_i$	(3) $S_r^2$	(4) $\hat{\rho}_i$	(5) $S_{e_{\rho_i}}^2$	(6) $\hat{\rho}$	(7) $E(S_{e_{\rho_i}}^2)$	(8) $S_{\rho}^2$	(9) $\sigma_{\rho}^2$	(10) ProVar	(11) $r_{\rho\rho}$	(12) $\hat{\rho}_{tr}$
<i>Achievement</i>												
Global Performance	26	.10	.01080	.20	.00166	.19	.00281	.00739	.00458	.38	.62	.20
Task Performance	26	.13	.01350	.25	.00192		(.053)	(.086)	(.068)			.23
Job Dedication	15	.20	.03260	.39	.00826							.31
Interpersonal Facilitation	18	.06	.01240	.11	.00232							.14
CWB (r)	13	.00	.01390	.00	.00962							.07
<i>Dependability</i>												
Global Performance	15	.13	.01400	.25	.00345	.29	.00492	.00803	.00311	.61	.39	.27
Task Performance	11	.09	.02290	.17	.00743		(.070)	(.090)	(.056)			.24
Job Dedication	7	.23	.01120	.46	.00640							.36
Interpersonal Facilitation	4	.11	.00590	.23	.00645							.27
CWB (r)	16	.21	.02170	.34	.00356							.31
<i>Order</i>												
Global Performance	26	.05	.01140	.10	.00175	.09	.00298	.00302	.00004	.99	.01	.09
Task Performance	26	.08	.02010	.16	.00309		(.055)	(.055)	(.007)			.09
Job Dedication	13	.05	.01160	.10	.00357							.09
Interpersonal Facilitation	18	-.01	.01640	-.02	.00364							.09
CWB (r)	13	.04	.01830	.07	.00431							.09
<i>Cautiousness</i>												
Global Performance	22	-.01	.01680	-.01	.00076	.04	.00203	.00322	.00119	.63	.37	.02
Task Performance	18	.06	.00890	.11	.00166		(.045)	(.057)	(.035)			.07
Job Dedication	6	.04	.01110	.08	.00740							.05
Interpersonal Facilitation	12	.00	.00880	.00	.00660							.03
CWB (r)	14	.06	.01570	.11	.00377							.07

Note. CWB = counterproductive work behavior. True score correlations for CWB are reverse coded (r). See notes to Table 2 for definitions of columns (1) through (12).

Orvis, Lebiecki, and Cortina (2006). Across the proportion of variance values shown in column 10 of Table 3, on average across the four facets of the Conscientiousness trait, 65% of the cross-criterion variance in true score correlation is explained by second order sampling error. In particular, the second order meta-analysis results show that almost all of the variability of the true score correlations for the order facet across job performance criteria is due to second order sampling error, indicating there is no differential validity by criterion type for this facet. For the dependability and cautiousness facets, more than 60% of the cross-criterion variance in true score correlation is explained by second order sampling error. For the achievement facet, about 40% of the cross-criterion variance is due to second order sampling error. The more accurate regression-based true score correlation estimates are reported in Column 12. Overall, these results suggest that the validities of all major Conscientiousness facets (with the possible exception of the achievement facet) do not differ much by criterion type; differences in job performance criterion are not likely to moderate validity for most Conscientiousness facets. This conclusion is contrary to that reached by Johnson et al. (2010), who suggested the presence of considerable differential validity by different job performance criteria.

In sum, second order meta-analysis is a useful tool to test differential validity hypotheses or whether validity meaningfully differs by moderators (e.g., different job performance criteria, ethnic groups, job characteristics).

*Case 3: moderator analysis particularly dealing with a moderator with more than two classes/conditions*

Mesmer-Magnus, DeChurch, Jimenez-Rodriguez, Wildman, and Shuffler (2011) examined whether the true score relationship between information team sharing and team performance varies across levels of three dimensions of team virtuality (e.g., extent of reliance on virtual tools, informational value, and synchronicity afforded by the tools). All meta-analytic estimates were based on the artifact distribution method and, more importantly, were completely statistically independent. Hence, first order meta-analytic

results for a given relationship fully met the assumptions of second order meta-analysis (see also the section below).

To determine the degree to which the differences in true score correlation across levels for a given dimension of team virtuality were due to second order sampling error, we conducted a second order meta-analysis for each dimension of team virtuality. Note that this second order meta-analysis was conducted using only the first order meta-analytic results reported in Mesmer-Magnus et al. (2011). The proportion of variance values shown in column 10 of Table 4 are all 100%. That is, all of the variance across varying levels of each team virtuality dimension is completely explained by second order sampling error. The more accurate regression-based true score correlation estimates are reported in Column 12. Overall, these results suggest that the true relationships between team information sharing and team performance do not differ by the level of team virtuality; differences in level of team virtuality are unlikely to moderate the true score correlation between team information sharing and team performance. This conclusion is starkly different from that reached by Mesmer-Magnus et al. (2011), who concluded that the degree of team virtuality “set[s] important boundary conditions for the information sharing – team performance relationship” (p. 221); i.e., Mesmer-Magnus et al. (2011) concluded that the relationship between team information sharing and team performance is stronger in hybrid teams (using both face to face and fully virtual interaction tools), teams using tools of moderate information values, and teams using a variety tools (moderate synchronicity). What we found through the second order meta-analyses as shown in Table 4 is that the relationship between team information sharing and team performance is consistent across teams regardless of the degree of each team virtuality dimension. That is, team virtuality does not set a boundary condition.

We note that in some first order meta-analyses, the subgroups are not statistically independent. For example, in a validity generalization meta-analysis, each sample in each primary study usually produces estimates of the validity of several abilities. So, for example, the subgroup results for verbal ability will not be independent of the subgroup results for quantitative ability. In the Mesmer-Magnus et al. (2011) meta-analyses used in our Case 3 illustrative

**Table 4**  
Second order meta-analysis of true score correlations of team information sharing with performance across dimensions of virtuality (Mesmer-Magnus et al., 2011).

Moderator Moderator Class	(1) <i>k</i>	(2) $\bar{r}_i$	(3) $S_r^2$	(4) $\hat{\rho}_i$	(5) $S_{e_{\rho_i}}^2$	(6) $\hat{\rho}$	(7) $E(S_{e_{\rho_i}}^2)$	(8) $S_{\hat{\rho}}^2$	(9) $\sigma_{\hat{\rho}}^2$	(10) ProVar	(11) $r_{\rho\rho}$	(12) $\hat{\rho}_{ir}$
<i>Use of virtual tools</i>												
None (Face to Face)	42	.34	.03240	.36	.00086	.37	.00193	.00174	.00000	1.00	.00	.37
Hybrid	8	.42	.02560	.46	.00384		(.044)	(.042)	(.000)			.37
Full (Fully Virtual)	6	.28	.03610	.31	.00738							.37
<i>Informational value</i>												
Low	3	.33	.00250	.34	.00088	.36	.00110	.00102	.00000	1.00	.00	.36
Moderate	9	.42	.03240	.46	.00432		(.033)	(.032)	(.000)			.36
High	44	.34	.02890	.36	.00074							.36
<i>Synchronicity</i>												
Moderate (Lagged)	8	.42	.02890	.46	.00433	.38	.00134	.00130	.00000	1.00	.00	.38
High (Real Time)	46	.34	.03240	.36	.00079		(.037)	(.036)	(.000)			.38

*Note.* Use of Virtual Tools indicates the proportion of team interaction that occurs via virtual tools – None (i.e., teams using only face to face interaction tools), Hybrid (i.e., teams using both face to face and virtual interaction tools), and Full (e.g., fully virtual teams only using virtual interaction tools); Informational value refers to the extent to which virtual tools transmit data that is valuable for team effectiveness; and Synchronicity indicates the extent to which team interactions occur in real time (e.g., phone and teleconferences) versus incurring a time lag (e.g., email) (Mesmer-Magnus et al., 2011). See notes to Table 2 for definitions of columns (1) through (12).

example all subgroup analyses were statistically independent of each other.

In sum, given statistical independence, second order meta-analysis is particularly useful in testing moderators that have more than two classes (or conditions) as shown above. If a moderator has only two classes, then moderator analysis can be conducted by checking the overlap in the 95% confidence intervals between the two first order meta-analytic mean estimates. However, when a moderator has more than two classes (e.g., the first two dimensions of team virtuality in Table 4), comparing the overlap in the 95% confidence intervals through the first order meta-analysis may not be a good solution because simultaneously comparing three or more 95% confidence intervals for overlap is not feasible in most cases. More importantly, this procedure does not reveal how much (if any) of the variance across first order meta-analytic mean estimates is explained by the moderator and how much is explained by second order sampling error. Finally, we note that second order meta-analysis can gauge not only whether or not a moderator exists, but also to what extent the moderator explains the variance in effect size for a given relationship. By contrast, comparing the 95% confidence intervals of first order meta-analytic means can answer only whether a moderator exists or not, but not to what extent the moderator explains the variance (differences) in effect size for a given relationship.

### Requirement of statistical independence of the *m* meta-analyses

The requirement of statistical independence means that the primary studies/samples contained in any first order meta-analysis should not also be included in any of the other first order meta-analyses. This requirement was clearly met in Cases 1 and 3, and relatively well in Case 2. For example, in Case 1, only studies conducted in China on Chinese subjects were included in the meta-analysis for China—and similarly for the other three countries. This assumption is important given that second order meta-analysis can be used to test a moderator (e.g., cross-country differences as a potential moderator of the personality-performance relationship in Case 1; differences in job performance criteria as a moderator of Conscientiousness facet validity in Case 2; and the degree of team virtuality as a potential moderator of the team information sharing and team performance relationship in Case 3). In testing such potential moderators, statistical independence should be met across classes of the moderator (across meta-analyses results for four East Asian countries in Case 1; across meta-analytic results for different

job performance criteria in Case 2; and across different degrees of team virtuality in Case 3). A minor violation as in Case 2 may affect the results in minor ways but will not change conclusions (Cooper & Koenka, 2012; Tracz et al., 1992). Nonetheless, we recommend that researchers do their best to meet this assumption of statistical independence to produce optimally accurate results when conducting second order meta-analysis. Finally, as noted earlier, this assumption also needs to be met in comparing first order meta-analytic subgroup mean estimates (and checking their 95% confidence intervals for overlap) as an approach to moderator analysis; that is, this assumption is not a limitation specific to second order meta-analysis.

Second order meta-analysis is appropriate for the case in which cumulative meta-analyses on the same question have been conducted with no or few overlapping studies across the meta-analyses over time. For example, suppose the first meta-analysis on a particular relation (e.g., the correlation between job satisfaction and job performance) was conducted in 1981, the second in 1991, and the third in 2001. If the three meta-analyses were conducted with no (or relatively few) overlapping studies (that is, the first including studies conducted until the end of 1980, the second including studies conducted between 1981 and 1990, and the third including studies conducted between 1991 and 2000), it is appropriate to conduct a cumulative second order meta-analysis to determine the degree of second order sampling error and other estimates (e.g., second order, grand mean estimates); in this case, time (or differences in decade) can be tested as a moderator and second order meta-analysis can serve as a tool for updating prior meta-analyses. However, if each meta-analysis after the first one contains all or most of the primary studies in the first meta-analysis plus a group of newer studies, then the different meta-analyses will not be statistically independent and it is inappropriate to conduct a second order meta-analysis.

In other cases, the statistical independence assumption will be met. For example, suppose the question is whether a particular training program or selection procedure works in the same way for several minority groups and the majority group. Separate meta-analyses may then be conducted for Blacks, Whites, Hispanic Americans, and Asian Americans. In such a case, the different meta-analyses would be expected to be statistically independent, because no individual should have been included in more than one of these sub-groups. That is, second order meta-analysis will be useful in research on differential validity (e.g., whether or to what extent the validity of cognitive tests will differ across ethnic groups). Another example in which the statistical independence assumption should be met would be second order meta-analysis

of validity of personality measures between Europe (e.g., Salgado, 1997, 1998) and North America (e.g., Barrick & Mount, 1991; Hurtz & Donovan, 2000).<sup>5</sup> Other examples of statistically independent meta-analyses examining the same relation were provided in an earlier section. The reader can probably think of other such scenarios.

It may prove to be the case that a major value of second order meta-analysis will be the analysis of cross-cultural and cross-national validity generalization studies. For example, recent meta-analysis study by Li and Cropanzano (2009) compared the relationships between justice perceptions and other attitudinal outcomes (e.g., trust, commitment, satisfaction) between North America and East Asia. They conducted East Asian meta-analyses and compared the meta-analytic results with corresponding, pre-existing North American meta-analytic results (Colquitt, Conlon, Wesson, Porter, & Ng, 2001). What they compared (see their Table 2, p. 797) was sample size weighted observed correlations and their 95% confidence intervals (CIs) for East Asia and North America. They concluded that there is no statistically significant difference between the two regions for some relationships (e.g., the procedural justice – turnover intention relationship) in that the two corresponding 95% CIs overlap. We believe that although well executed, the Li and Cropanzano (2009) meta-analysis would have been more informative if they had used the second order meta-analysis described in this paper using corrected correlations. (In fairness to the authors, we note that at the time they conducted their meta-analysis second order meta-analysis methods were not available.) Using second order meta-analysis, they could have provided the *degree*, rather than *dichotomy*, of true cross-cultural differences (=1 minus the proportion of cross-region variance due to second order sampling error) and some of the other important information shown in our Tables 2–4 (e.g., second order grand mean estimates and regressed estimates for country mean effect sizes).

Although it may go without saying, first order meta-analytic studies included in second order meta-analyses should all examine the same relationship, using ostensibly comparable measures and samples (e.g., self-report measures of Big Five traits, supervisor ratings of performance, employees), and comparable meta-analytic methods (see Spector, 2001, p. 23 for more details). Finally, all first order meta-analytic studies included in the second order meta-analyses should be of high quality and transparency and free of publication bias (Aytug, Rothstein, Zhou, & Kern, 2012; Kepes, Banks, McDaniel, & Whetzel, 2012). Ideally, the question of possible publication bias should be satisfactorily addressed in each first order meta-analysis, and this requirement is increasingly being imposed on published meta-analyses. However, where this has not been done, the researcher applying second order meta-analysis will need to consider the possibility of publication bias and may need to contact the authors of the first order meta-analysis to make this determination. The accuracy of the second order meta-analyses depends on the accuracy of the first order meta-analyses synthesized, whose accuracy in turn depends on the accuracy of the primary studies synthesized.

<sup>5</sup> In fact, we compared Hurtz and Donovan (2000) and Salgado (1998), both of which examined the criterion-related validity of self-report Big Five personality measures for predicting supervisor ratings of job performance in North America and Europe, respectively. Overall, our second order meta-analysis results clearly suggest that the operational validities of the Big Five personality measures (except for Agreeableness) are generalizable between North America and Europe; cross-regional differences are unlikely to moderate validity for most Big Five personality traits. It is interesting that Agreeableness does have a close to zero relationship with job performance in Europe, but not in North America. It is noted that we considered using Barrick and Mount (1991) instead of Hurtz and Donovan (2000), but unfortunately, Barrick and Mount (1991) did not report the observed variance (and SD) across validities, a necessary input to the second order meta-analysis, so we had to use Hurtz and Donovan (2000) instead. It is also noted that Salgado (1998) is an update of Salgado (1997) with larger numbers of validation studies. Detailed results are displayed in the Appendix A.

## Reporting standards for first order meta-analyses

In our presentation of the methods of second order meta-analysis, we defined the items of information from the first order meta-analyses that are needed for application of second order meta-analysis methods. For the sake of future cumulative research knowledge, it is critical that this information be reported. For bare bones second order meta-analyses, these include the variance of the observed correlations (which is sometimes not reported) and mean observed correlation (almost always reported). These are also the values that are critical for artifact distribution meta-analyses. When the first order meta-analyses have corrected each correlation individually, it is critical to report not only the mean corrected correlation (always reported), but also the variance of the corrected correlations (not always reported). Other values (such as *k*, the number of studies in the meta-analysis and the mean corrected correlation) are virtually always reported.

## Limitations and potential criticisms of second order meta-analysis

One limitation of second order meta-analysis methods is that the requirement for statistical independence of meta-analysis could limit the frequency with which the methods can be applied. The extent to which moderate violations of this assumption affect the results is unknown. Cooper and Koenka (2012), in discussing an older, cruder form of second order meta-analysis (our option 2, discussed earlier) suggest that minimizing the lack of independence might be sufficient to produce reasonably accurate results and they give several examples of such published second order meta-analyses. And, as noted earlier, Tracz et al. (1992) found that violations of independence had almost no effect on the results of first order meta-analyses. In general, we believe this is a question that will require further research to produce a definitive answer; more studies similar to the Tracz et al.'s (1992) study are needed. But even if it turns out that second order meta-analysis cannot be applied on a broad basis, it is important to note that a number of other procedures with limitations on frequency of use are well represented in the literature. These include methods for latent growth modeling (e.g., analysis of longitudinal data collected over multiple time periods), generalizability theory of reliability, and a number of complex structural equation modeling (SEM) methods (e.g., multi-level SEM). In each case, the procedures are quite useful when they can be applied and this is viewed in the methodological literature as compensating for the fact that they cannot be applied frequently.

Second order meta-analysis is not directly concerned with the variability of study population correlations *within* each of the first order individual meta-analyses. To be sure, this within-meta-analysis variability (i.e., non-artifactual variability between-primary studies in first order meta-analyses) is taken into account mathematically in second order meta-analysis methods, as can be seen in Eqs. (5a)–(5c), (6a)–(6c), 7a, 7b, and 7c. However, a finding that second order sampling error accounts for all of the variability in the mean values across first order meta-analyses does not imply that population parameters do not vary within first order meta-analyses. Such a finding simply means that the *mean* values are equal across the different first order meta-analyses. For example, our finding that the mean meta-analytic operational validity for Conscientiousness is the same across four East Asian countries does not mean that this validity cannot vary somewhat across sub-populations within, say, within South Korea. If this is the case, this variability will be reflected in the results of the first order meta-analysis. It is the purpose of the original first order meta-analyses to address this non-artifactual between-primary-studies variability within each first order meta-analytic context. The purpose of

second order meta-analysis is to gauge the true (i.e., non-artificial) between-meta-analyses (e.g., cross-country, cross-region, cross-criterion, cross-setting) variability of mean effect size values of ostensibly the same relationship and to use this information to improve accuracy of estimation for each first order meta-analytic mean estimate.

A possible objection to second order meta-analysis is the following: Instead of second order meta-analysis, why not conduct an overall meta-analysis pooling all primary study data from all meta-analyses (which will yield the same grand mean as the second order meta-analysis), and then break out into sub-meta-analyses based on hypothesized moderators (which yields the same subgroup means as those used in the second order meta-analysis)? First, this is often an impossible or impractical alternative. For example, in all cases illustrated above, the primary studies used in all first order meta-analyses were not available. Some journals (e.g., *Journal of Applied Psychology*) in the fields of Organizational Behavior and Human Resource Management have only recently required that meta-analysts be transparent about their procedures and report all data from primary studies used in their meta-analysis (Aytug et al., 2012; Kepes et al., 2012). As mentioned, second order meta-analysis can be conducted using only first order meta-analytic results ( $k$ , mean observed  $r$ , mean corrected  $r$ , and variance across observed or corrected  $r$ s) and thus it can be applied to most if not all previous first order meta-analyses. Second, such omnibus meta-analyses typically have problems of lack of independence of subgroup meta-analyses. For example, in a typical validity generalization meta-analysis each primary study estimates validity for several ability measures. So the separate subgroup analyses for, say, verbal and quantitative ability are not independent. Third, and perhaps most important, this procedure does not allow one to estimate the variance (and the percentage of variance) across sub-group meta-analyses that is (and is not) due to second order sampling error variance, because second order sampling error variance is not computed (or computable) in the omnibus meta-analysis approach. This is because omnibus meta-analyses and their sub-group meta-analyses are both first order meta-analyses. For example, application of this approach to the Conscientiousness validity data in our first example would not have revealed that all the variance across the four East Asian countries in meta-analytic operational validity values was due to second order sampling error. Instead, the values would have been taken at face value. Likewise, it would not have been possible to improve the accuracy of the estimates of meta-analytic operational validity for the other Big Five measures for which not all the variance across countries was accounted for by second order sampling error. That is, the analysis shown in Tables 2–4 could not be performed. So the omnibus meta-analysis procedure is not a substitute for second order meta-analysis.

A variation on this objection is the following: Why not just conduct an omnibus, pooled meta-analysis along with sub-group meta-analyses based on hypothesized moderators and then look at the relative variances? The difference between the estimated population parameter variance in the omnibus meta-analysis and the average of this figure across the sub-group meta-analyses estimates the variance of the subgroup means (the variance of means across subgroup meta-analyses). This statement reflects the well known analysis of variance (ANOVA) principle that total variance is the sum of between group variance and average within group variance. However, knowing the variance of the subgroup means does not allow one to estimate how much of this variance is (or is not) due to second order sampling error and therefore does not allow computation of the proportion of this variance that is due to second order sampling error. As a result, the analyses presented in our example cannot be conducted. For example, if all the between-mean variance was accounted for by second order

sampling error (as was the case with Conscientiousness in our first example application), there would be no way for one to know this. The procedure advocated here allows one to compute the percentage of *total variance* that is accounted for by between group variance in mean values, but this is not the same as the percentage of between-group variance in mean values that is due to second order sampling error variance. So again, this is a procedure that is not a substitute for second order meta-analysis.

Related to the aforementioned ANOVA principle, researchers may want to know whether the true (non-artificial) between-group variance in mean (meta-analytic) values across subgroup meta-analyses is larger or smaller than the mean true (non-artificial) within-group variance across the same subgroup meta-analyses. This is computed as the ratio of true between-group (i.e., between-meta-analyses) variance to average true within-group (i.e., within-meta-analyses) variance:

$$\text{Variance ratio for psychometric meta-analyses} = \hat{\sigma}_{\rho}^2 / E(S_{\rho}^2),$$

where the numerator is the estimated true between-meta-analyses variance in mean (meta-analytic) values across  $m$  first order meta-analyses (see Eqs. (6) and (7)) and the denominator is the weighted mean of the estimated true within-meta-analysis variance across  $m$  first order meta-analyses (see Eqs. (3) and (4)). Because most first order psychometric meta-analyses report true variance in the form of standard deviation ( $SD_{\rho}$ ), this ratio can be computed in most cases. Given many readers' familiarity with the ANOVA principle, we have to caution against using this variance ratio (which is analogous to the  $F$  static) to determine statistical significance. This is problematic particularly because sample size ( $m$ ) in this context is the number of the first order meta-analyses contributing to a second order meta-analysis. Further, we note that, like first order meta-analysis methods, second order meta-analysis methods presented here are also developed to overcome problems in statistical significance testing.

Another possible objection is this: Why not just compute a meta-regression in which coded hypothesized moderators are used to predict the primary study correlations pooled across all the first order meta-analyses? (These correlations can be either observed correlations, as in bare bones meta-analysis, or correlations corrected for measurement error.) This procedure fails for the same reason as above: The squared multiple correlation (after the appropriate adjustment for capitalization on chance) will reveal the percentage of the total variance that is accounted for by the hypothesized moderator or moderators. But it will not reveal the percentage of the variance in the mean values that is explained by second order sampling error, and therefore the analyses allowed by second order meta-analysis cannot be done. That is, one could again not obtain the information (e.g., true between-meta-analysis variance) presented in our examples in Tables 2–4. So this procedure is also not capable of being a substitute for second order meta-analysis.

## Conclusion

In conclusion, the methods of second order meta-analysis provide unique information that cannot be obtained using the more traditional methods of first order meta-analysis. The methods are particularly useful in conducting cross-culture generalization studies (i.e., synthesizing first order meta-analyses conducted in different countries for the same relationship) and meta-analytic moderator analyses (i.e., comparing first order meta-analytic results of the same relationship across different settings and/or groups). This unique information can be important from the point of view of cumulative knowledge and understanding, as illustrated in our empirical examples.

**Appendix A.**

Second order meta-analysis of operational validities of the Big Five personality measures between North America (Hurtz & Donovan, 2000) and Europe (Salgado, 1998).

Predictor	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Region	<i>k</i>	$\bar{r}_i$	$S_r^2$	$\hat{\rho}_i$	$S_{e_{\hat{\rho}_i}}^2$	$\hat{\rho}$	$E(S_{e_{\hat{\rho}_i}}^2)$	$S_{\hat{\rho}}^2$	$\sigma_{\hat{\rho}}^2$	ProVar	$r_{\rho\rho}$	$\hat{\rho}_{ir}$
<i>Conscientiousness</i>												
North America	45	.14	.01610	.20	.00073	.16	.00068	.00159	.00091	.43	.57	.18
Europe	35	.09	.01260	.12	.00064		(.026)	(.040)	(.030)			.14
<i>Emotional Stability</i>												
North America	37	.09	.00840	.13	.00047	.14	.00046	.00002	.00000	1.00	.00	.14
Europe	49	.10	.01120	.14	.00045		(.021)	(.005)	(.000)			.14
<i>Agreeableness</i>												
North America	40	.07	.01080	.11	.00067	.03	.00032	.00181	.00150	.17	.83	.10
Europe	37	.01	.00770	.01	.00021		(.018)	(.043)	(.039)			.01
<i>Extraversion</i>												
North America	39	.06	.01110	.09	.00064	.09	.00067	.00000	.00000	1.00	.00	.09
Europe	45	.07	.01900	.09	.00070		(.026)	(.000)	(.000)			.09
<i>Openness</i>												
North America	35	.04	.00930	.06	.00060	.05	.00053	.00002	.00000	1.00	.00	.05
Europe	28	.04	.00850	.05	.00047		(.023)	(.005)	(.000)			.05

Note: Columns (1) through (4) are input values (italicized) available from first order meta-analyses. (1) Number of samples; (2) Sample size weighted mean observed correlations; (3) Sample size weighted observed variance across observed correlations; (4) First order meta-analytic mean true-score correlation estimates; (5) Second order sampling error variance for each first order meta-analytic true score correlation estimate (see discussion of Eq. (7d)); (6) Second order, grand mean true score correlation estimates (Eq. (7b)); (7) Expected (average) second order sampling error variance (Eq. (7d)) and standard error (in parentheses); (8) Observed variance and SD (in parentheses) across first order mean true score correlation estimates (Eqs. (7a), (7b), and (7c)); (9) Estimated true variance and SD (in parentheses) across first order mean true score correlation estimates after expected second order sampling error variance is subtracted out from the observed variance (Eq. (7)); negative values are set to zero; (10) The proportion (percentage if multiplied by 100) of the observed variance across first order mean true score correlation estimates that is due to second order sampling error variance; values greater than 1 are set to 1; (11) The reliability of the first order meta-analytic true score correlation vectors; these values are computed as 1 minus the values in Column 10; (12) Regressed first order true score correlation estimates based on the reliability of the original true score correlation vectors shown in Column 11.

**References**

Aguinis, H., Gottfredson, R. K., & Wright, T. A. (2011). Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior*, 32, 1033–1043.

Aytug, Z. G., Rothstein, H. R., Zhou, W., & Kern, M. C. (2012). Revealed or concealed? Transparency of procedures, decisions, and judgment calls in meta-analyses. *Organizational Research Methods*, 15, 103–133.

Bangert-Drowns, R. L. (1986). Review of developments in meta-analysis methods. *Psychological Bulletin*, 99, 388–399.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection & Assessment*, 9, 9–30.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Subgroup analyses. Introduction to meta-analysis*. Chichester, UK: Wiley.

Brannick, M. T., Yang, L., & Cafri, G. (2011). Comparison of weights for meta-analysis of *r* and *d* under realistic conditions. *Organizational Research Methods*, 14, 587–607.

Carlson, K. D., & Ji, F. X. (2011). Citing and building on meta-analytic findings: A review and recommendations. *Organizational Research Methods*, 14, 696–717.

Chan, M. E., & Arvey, R. D. (2012). Meta-analysis and the development of knowledge. *Perspectives on Psychological Science*, 7, 79–92.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, 86, 425–445.

Combs, J. G., Liu, Y., Hall, A. T., & Ketchen, D. J. (2006). How much do high performance work practices matter? A meta-analysis of their effects on organizational performance. *Personnel Psychology*, 59, 501–528.

Cooper, H., & Koenka, A. C. (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist*, 67, 446–462.

DeGeest, D., & Schmidt, F. L. (2010). The impact of research synthesis methods on industrial-organizational psychology: The road from pessimism to optimism about cumulative knowledge. *Research Synthesis Methods*, 1, 185–197.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of Conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validities of narrow traits. *Journal of Applied Psychology*, 91, 40–57.

Ghiselli, E. E. (1964). *Theory of psychological measurement*. New York: McGraw-Hill.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–300). New York: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London: Academic Press.

Hofstede, G. (2010). The GLOBE debate: Back to relevance. *Journal of International Business Studies*, 41, 1339–1346.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed vs. random effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment*, 8, 275–292.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage.

Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85, 869–879.

Johnson, J. W., Steel, P., Scherbaum, C. A., Hoffman, C. C., Jeanneret, P. R., & Foster, J. (2010). Validation is like motor oil: Synthetic is better. *Industrial and Organizational Psychology*, 3, 305–328.

Kepes, S., Banks, G. C., McDaniel, M. A., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods*, 15, 624–662.

Li, A., & Cropanzano, R. (2009). Do East Asians respond more/less strongly to organizational justice than North Americans? A meta-analysis. *Journal of Management Studies*, 46, 787–805.

Magnusson, D. (1966). *Test theory*. New York: Addison-Wesley.

McDaniel, M. A. (2007). Validity generalization as a test validation approach. In S. M. McPhail (Ed.), *Alternative validation strategies* (pp. 159–180). San Francisco: Jossey Bass.

Mesmer-Magnus, J. R., DeChurch, L. A., Jimenez, M. J., Wildman, J., & Shuffler, M. (2011). A meta-analytic examination of virtuality and information sharing in teams. *Organizational Behavior and Human Decision Processes*, 115, 214–225.

Murphy, K. R., & Newman, D. A. (2003). The past, present, and future of validity generalization. In K. R. Murphy (Ed.), *Validity generalization* (pp. 403–424). Mahwah, NJ: Erlbaum Associates.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Oh, I.-S. (2009). *The Five-Factor Model of personality and job performance in East Asia: A cross-cultural validity generalization study*. Unpublished doctoral dissertation. Iowa City, IA: University of Iowa.

Podsakoff, P. M., Bommer, W. H., Podsakoff, N. P., & MacKenzie, S. B. (2006). Relationships between leader reward and punishment behavior and

- subordinate attitudes, perceptions, and behaviors: A meta-analytic review of existing and new research. *Organizational Behavior and Human Decision Processes*, 99, 113–142.
- Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–322). New York: Russell Sage Foundation.
- Raudenbush, S. W. (2009). Analyzing effect sizes: Random effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 295–316). NY: Russell Sage.
- Sackett, P. R. (2003). The status of validity generalization research: Key issues in drawing inferences from cumulative research findings. In K. R. Murphy (Ed.), *Validity generalization* (pp. 91–114). Mahwah, NJ: Erlbaum Associates.
- Salgado, J. F. (1997). The five-factor model of personality and job performance in the European community. *Journal of Applied Psychology*, 82, 30–43.
- Salgado, J. F. (1998). The Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance*, 11, 271–288.
- Schmidt, F. L. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods*, 11, 96–113.
- Schmidt, F. L. (2010). Detecting and correcting the lies that data tell. *Perspectives on Psychological Science*, 5, 233–242.
- Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed vs. random models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, 62, 97–128.
- Schmidt, F. L., Shaffer, J. A., & Oh, I.-S. (2008). Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology*, 61, 827–868.
- Spector, P. E. (2001). Research methods in industrial and organizational psychology: Data collection and data analysis with special consideration to international issues. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 10–26). Thousand Oaks, CA: Sage.
- Tracz, S. M., Elmore, P. B., & Pohlmann, J. T. (1992). Correlational meta-analyses: Independent and nonindependent cases. *Educational and Psychological Measurement*, 52, 879–888.
- Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, 96, 1167–1194.
- Wang, G., Oh, I.-S., Courtright, S. H., & Colbert, A. E. (2011). Transformational leadership and performance across criteria and levels: A meta-analytic review of 25 years of research. *Group & Organizational Management*, 36, 223–270.