

# Estimating Group-Level Relationships: General Recommendations and Considerations for the Use of Intraclass Correlation Coefficients

Maura I. Burke<sup>1</sup> · Ronald S. Landis<sup>1</sup> · Michael J. Burke<sup>2</sup>

Published online: 26 July 2016  
© Springer Science+Business Media New York 2016

## Abstract

**Purpose** In this investigation, we argue for why and how available intraclass correlation coefficients and other types of reliability estimates can be employed as sample-based reliability estimates within primary and meta-analytic studies when relationships between group-level phenomena are of interest.

**Design/Methodology/Approach** Group-level correlations and reliability estimates were obtained from 46 studies examining organizational climate–performance relationships. We illustrate how the group-level reliability estimates can be used to correct correlations for predictor and criterion unreliability. Procedures are presented for computing the sampling variances of individually corrected correlations that account for sampling error in the group-level reliability estimates.

**Findings** Support was found for the conservative nature of meta-analytic parameter estimates when group-level reliability information is sample-based as opposed to assumed population values. In addition, our analyses indicated that

conclusions about substantive relationships between group-level variables can change based on availability of sample-based reliabilities within both primary and meta-analytic studies.

**Implications** Results from this study suggest that researchers should rely on sample-based meta-analytic procedures when examining the generalizability of group-level relationships. This study also demonstrates the importance of using all available reliability information and accounting for sampling error in the reliability estimates when conducting meta-analyses at the group level of analysis.

**Originality/Value** This study breaks ground by systematically examining the use of intraclass correlation coefficients as reliability estimates within group-level meta-analytic studies. Furthermore, illustrative analyses provide guidance to primary and meta-analytic researchers in regard to how to correct group-level correlations for unreliability in the predictor, criterion, or both whenever and in whatever proportions the artifact information is available.

**Keywords** Meta-analysis · Group-level data · ICCs · Reliability

The RBNL meta-analytic program is available upon request from the third author at mburke1@tulane.edu.

✉ Maura I. Burke  
mburke6@hawk.iit.edu

Ronald S. Landis  
rlandis@iit.edu

Michael J. Burke  
mburke1@tulane.edu

<sup>1</sup> Department of Psychology, Illinois Institute of Technology, 3105 S. Dearborn St., LS 252, Chicago, IL 60616, USA

<sup>2</sup> A.B. Freeman School of Business, Tulane University, Goldring/Woldenberg Hall, 7 McAlister Drive, New Orleans, LA 70118, USA

## Introduction

Meta-analysis is a statistical procedure widely used by researchers to cumulate results from primary studies with the overall goal of drawing accurate inferences about relationships between constructs. Although meta-analysis is frequently conducted with data gathered at the individual level of analysis, relatively few studies have employed meta-analytic procedures with aggregated data to estimate relationships at the group, or business unit, level of analysis. Notable examples of group-level meta-analyses

include Harter et al. (2002, 2010), Wallace et al. (2016), and Whitman et al. (2010, 2012). In most cases, group-level meta-analytic studies either have not corrected for measurement error in the predictor and criterion measures or have relied on assumed population predictor and criterion reliability information. As discussed by Raju et al. (1991), these forms of meta-analysis are not optimal when predictor and criterion reliability information is available from primary studies.

The use of suboptimal meta-analytic procedures is clearly understandable when it is the case that reliability information is unavailable. We contend, however, that sample-based reliability in studies involving aggregated data may be available in the form of reported intraclass correlation coefficients (ICCs). At present, primary studies and published meta-analyses have not used this sample-based ICC information to correct group-level correlations for measurement error in the predictor or criterion measure and to account for sampling error in such reliability estimates. The current paper illustrates the logic and method by which available ICCs can be employed as reliability estimates within primary and meta-analytic studies when group-level phenomena are of interest. In addition, this paper provides illustrative comparisons of meta-analytic scenarios where different *types* of group-level reliabilities (e.g., sample-based ICCs and literature-based, assumed population reliability estimates) are incorporated into tests of three research questions. We offer guidance on how to handle missing group-level reliabilities, and further, clarify how primary and meta-analytic study conclusions are affected by the type of reliability value used, sample-based versus assumed population, when such information is missing.

The structure of the current paper is as follows. First, examples of prior meta-analytic approaches using aggregated data and explanations concerning how reliability information was handled in these cases are identified and discussed. Following, a case is made for using available ICC information from primary studies as estimates of group-level predictor and criterion reliability. We then illustrate how ICC information can be used to correct correlations for unreliability in primary studies and for estimating the sampling variances (and standard errors) of individually corrected correlations. Sampling variance equations for individually corrected correlations, covering all possible situations related to the availability of reliability information, are presented in conjunction with the assumptions associated with each situation. Finally, we demonstrate how sample-based meta-analyses can be conducted when such predictor and criterion reliability information is sporadically available. Here, we contrast results from sample-based meta-analyses with findings from meta-analyses based on the use of assumed

population reliabilities obtained from the literature. In addition, we present supplemental meta-analyses that artificially restrict the availability of reliability to further demonstrate how reductions in sample-based reliabilities affect meta-analytic parameter estimates.

### Previous Meta-Analytic Approaches Using Aggregated Data

The organizational research literature is replete with meta-analyses conducted at the individual level of analysis. Markedly fewer studies, however, have applied meta-analytic procedures to data gathered at the group or organizational level of analysis. This is unfortunate given that conducting meta-analysis with aggregated data is important for understanding and making generalizations about group- and organizational-level phenomenon. Of the studies that have employed meta-analytic procedures with group-level data, most have treated reliability estimates as assumed, population-based reliabilities (as opposed to sample-based reliability values). For instance, Harter et al. (2010) examined the causal impact of employee work perceptions on an organization's bottom line and corrected correlations for measurement error based on aggregated test-retest reliability estimates. In another study by Harter et al. (2002), similar correction procedures using test-retest reliabilities were employed in examining relationships between employee satisfaction, employee engagement, and business outcomes. In each of the preceding meta-analyses, sampling error in the reliability estimates was not taken into account.

Researchers have also included ICC(2) values when estimating the reliability of group-level variables in meta-analyses. For example, in a study examining relationships between satisfaction, citizenship behaviors, and performance in work units, Whitman et al. (2010) corrected correlations for unreliability using ICC(2) values. Although these values were sample-based, Whitman et al. employed meta-analytic procedures that treated the sample-based ICC(2) values as assumed, population values. Whitman et al. (2012) applied the same meta-analytic procedures. Interestingly, Hong et al. (2013) corrected observed correlations for different types of sample-based reliability estimates including ICC(2)s, but they did not employ meta-analytic procedures that account for sampling error in the sample-based reliabilities.

While the above studies certainly advance our understanding of relationships between group-level variables, the specific treatment and assumptions concerning reliability estimates that were employed in most of these studies are not optimal. Relying on assumed reliability values can produce inaccurate results because the degree to which assumed values mirror those of the population is often

unknown (Raju et al. 1989). Related, when using sample-based reliability values, failing to account for sampling error in reliability estimates can lead to imprecision in the estimates of the mean and variance of corrected correlations (see Raju et al. 1991).

In sum, meta-analysts have corrected group-level correlations for unreliability using ICC(2) information. Despite this, however, none of these studies have accounted for sampling error in the sample-based reliability estimates. The current study departs from those described above by demonstrating how group-level meta-analyses can be improved by using all available reliability information from the primary group-level studies and at the same time take into account sampling error in the reliabilities reported within the primary studies.

### ICCs as a Form of Group-Level Reliability

Two variations of ICCs, ICC(1) and ICC(2), are frequently used in organizational research. These coefficients are more generally interpreted as measures of the proportion of variance attributable to objects of measurement (McGraw and Wong 1996). The ICC(2) provides an estimate of the reliability of group means (Bartko 1976; Bliese 2000; Shrout and Fleiss 1979) and is calculated using the following formula<sup>1</sup>:

$$\text{ICC}(2) = \frac{\text{MS}_B - \text{MS}_W}{\text{MS}_B} \quad (1)$$

where  $\text{MS}_B$  is the mean square between groups and  $\text{MS}_W$  is the mean square within groups.

ICC(2) values share several important features with traditional reliability indices. Reliability is an index of the variance of interest over the sum of the variance of interest plus error (Shrout and Fleiss 1979). Consistent with the preceding definition, ICCs have long been recognized as reliability estimates (Cronbach et al. 1972; Ebel 1951; Lahey et al. 1983). For ICCs, when within-subject variance is small relative to total variance, the coefficient is larger than when within-subject variance is relatively large. Importantly, this is also true of reliability coefficients (e.g., coefficient alpha) commonly used in meta-analysis conducted on individual-level phenomena.

From the preceding points, ICC(2) estimates are frequently considered indices of reliability for group-level data (Cronbach et al. 1972; Ebel 1951; Lahey et al. 1983; Stanley 1971) and are expressed similarly to traditional reliability indices (i.e., correlation coefficients). Bartko (1976) originally termed ICCs as intraclass correlation reliability coefficients and “use of the intraclass correlation

(ICC) as an index of the reliability of ratings has been well documented and accepted in psychological research” (Lahey et al. 1983, p. 586). Most importantly, as discussed above, ICC(2)s are interpreted as reliability values in the organizational psychology literature (Hong et al. 2013; Whitman et al. 2010, 2012). Therefore, ICC(2) values may logically be used to correct observed, group-level correlations for measurement error.

Given ICC(2)s as estimates of group-level predictor and criterion reliability, the general equation for estimating the sampling variance of a corrected correlation (see Raju and Brand 2003; Raju et al. 1991) can be adjusted for use with ICC(2) values and other types of group-level reliability estimates (e.g., stability coefficients) whenever they are available for the predictor or criterion. We next discuss these equations and how they might be employed whenever group-level reliability information is available in primary studies.

### Correcting Correlations for Sample-Based Unreliability

Despite the long-standing availability of procedures to correct correlations within primary studies for unreliability with appropriately defined standard errors (see Raju et al. 1991), there has not been a systematic application of such procedures using available ICC values. This section presents the relevant formulas for correcting a correlation for predictor and criterion unreliability, as well as the general formula and special cases for estimating the sampling variances of individually corrected correlations. We represent equations from Raju et al. (1991, 2004) to illustrate their relevance to group-level studies and for corrections to correlations with ICC values and other group-level reliability estimates.

We begin by letting  $r_{xy}$  represent the restricted and attenuated effect between the predictor ( $x$ ) and criterion ( $y$ ) in a sample, where  $r_{xx}$  and  $r_{yy}$  represent the sample-based predictor and criterion reliability values, respectively. In addition, the range restriction factor  $k$  is defined as  $1/u$  where  $u$  is the ratio of the unattenuated, restricted standard deviation on  $x$  to the unattenuated, unrestricted standard deviation on  $x$  (Raju and Brand 2003). With the application of classical test theory (Lord and Novick 1968), an estimate of an unrestricted and unattenuated population correlation ( $\rho_{xy}$ ) can be written as:

$$\hat{\rho}_{xy} = \frac{kr_{xy}}{\sqrt{r_{xx}r_{yy} - r_{xy}^2 + k^2r_{xy}^2}} \quad (2)$$

When all sample-based artifact information is available, the general sampling variance formula associated with the corrected correlation as presented in Raju and Brand is:

<sup>1</sup> Equations used have been reported elsewhere as noted. Numbering of the equations in the current paper is for ease of presentation only.

$$\hat{V}(\hat{\rho}_{xy}) = \frac{k^2 r_{xx} r_{yy} (r_{xx} - r_{xy}^2)(r_{yy} - r_{xy}^2)}{(n-1)\hat{W}^3} \quad (3)$$

where

$$\hat{W} = r_{xx} r_{yy} - r_{xy}^2 + k^2 r_{xy}^2 \quad (4)$$

As discussed below, Eq. 3 can be adjusted depending on the availability of group-level reliability information.

### Availability of Group-Level Reliability Information

Individual studies vary in whether reliability information is provided for group-level variables. For instance, some studies report reliability (e.g., ICCs) information for the predictor (Chen et al. 2007; Dawson et al. 2008; Schmit and Allscheid 1995), criterion (Baer and Frese 2003), or both (Dietz et al. 2004; Salanova et al. 2005; Simons and Roberson 2003). In this section, we present sampling variance formulas, originally derived by Raju and Brand (2003), which apply to all possible situations concerning the availability of predictor and criterion reliabilities. In doing so, the assumptions and equations that allow for disattenuating group-level correlations and for conducting meta-analyses with sample-based reliabilities at the group-level of analysis are discussed.

The variance formulas presented below take into account sampling variance associated with  $r_{xx}$ ,  $r_{yy}$ ,  $r_{xy}$ , and their intercorrelations. It is important to note that for each formula,  $k$  will be fixed at one because we are not concerned with corrections for range restriction in the present discussion. As a result,  $W$  (i.e., Eq. 4) reduces to the product of  $r_{xx}$  and  $r_{yy}$ .

Clarification of several conceptual terms is important before presenting the sampling variance formulas for special cases of Eq. 3. The term “assumption” relates to the treatment of reliability information based on its availability within a primary study. That is, when predictor or criterion reliability information is missing, the reliability value that is employed is treated as “assumed fixed,” meaning the reliability coefficient is treated as a parameter not having sampling error. In this case, the average reliability from the set of primary studies or a reliability value from other investigations can be used as the assumed reliability. In most cases, meta-analysts rely on an assumed population reliability obtained from the literature such as the frequently employed criterion reliability value of .52 (see LeBreton et al. 2014). In contrast, the term “sample-based” means the reliability estimate is reported in the primary study and is, therefore, treated as a parameter with known sampling error.

When the criterion reliability is missing, it is assumed fixed, and the sampling variance formula can be written as:

$$\hat{V}(\hat{\rho}_{xy}) = \frac{r_{xx} r_{yy}^2 (r_{xx} - r_{xy}^2)(1 - r_{xy}^2)}{(n-1)\hat{W}^3} \quad (5)$$

An assumed reliability value would need to be used for  $r_{yy}$  in Eq. 5. In this case, a criterion reliability coefficient from prior investigations may be used as an estimate (Raju and Brand 2003), such as the .52 value noted above.

When the predictor reliability (e.g., ICC) is missing, it is assumed fixed. The sampling variance formula can be written as:

$$\hat{V}(\hat{\rho}_{xy}) = \frac{r_{xx}^2 r_{yy} (1 - r_{xy}^2)(r_{yy} - r_{xy}^2)}{(n-1)\hat{W}^3} \quad (6)$$

An assumed reliability value would need to be used for  $r_{xx}$  in Eq. 6. In this case, a predictor reliability coefficient from prior investigations (e.g., .80) may be used as an estimate.

When both criterion and predictor reliabilities are missing from a primary study, they are assumed fixed, and the sampling variance formula for the corrected correlation can be written as:

$$\hat{V}(\hat{\rho}_{xy}) = \frac{r_{xx}^2 r_{yy}^2 (1 - r_{xy}^2)^2}{(n-1)\hat{W}^3} \quad (7)$$

Assumed reliability values are required for  $r_{xx}$  and  $r_{yy}$  in Eq. 7. In this case, predictor and criterion reliability coefficients from prior investigations may be used as estimates (e.g., .80 and .52, respectively). The treatment of reliability information and sampling error in this situation is somewhat similar to how reliability and sampling error are handled within current group-level meta-analytic studies. That is, sampling error in the sample-based reliabilities is NOT taken into account.

The above sampling variance formulas cover all situations related to the availability of reliability information reported in primary studies. As noted in a previous section, ICC(2) values can logically be inserted for  $r_{xx}$  and  $r_{yy}$  whenever this reliability information is presented for the group-level predictor, criterion, or both. When ICC(2) values are inserted into the equations, based on assumptions associated with the availability of sample-based reliability information, the sampling variance formulas account for sampling error in the group-level reliability estimates.

We next illustrate how the preceding equations can be employed, with sporadically available reliability information, as part of sample-based meta-analyses to estimate the mean and variance of corrected correlations at the group level. As referenced within the literature, these procedures will be collectively referred to as the Raju–Burke–Normand–Langlois (RBNL) meta-analytic

procedures.<sup>2</sup> When applying these procedures, we also illustrate how individually corrected correlations and their sampling variances as well as how meta-analytic findings are affected by the use of available ICCs and other group-level reliability values versus the use of assumed population reliability values.

## Illustrative Examples

In an effort to provide realistic, illustrative examples of scenarios concerning both the availability of reliability information and the use of literature-based (assumed population) reliabilities for missing reliabilities, we focus on the organizational climate–work group performance literature. We do so for two reasons. First, the organizational climate domain has the largest number of published studies with aggregated group-level data within the fields of organizational psychology and management (Wallace et al. 2016). Second, performance is one of the more important criterion variables of interest in organizational psychology (Austin and Villanova 1992). For purposes of illustrating the preceding procedures with varying amounts of reliability information reported in primary studies, we focus on three distinct components of the criterion domain: behavioral-oriented, productivity-oriented, and health/safety outcomes. Based on past research (see Christian et al. 2009), we expect a climate focused on concern for employees to relate positively with behavior-oriented performance measures (RQ1) and productivity-oriented outcomes (RQ2), and we expect concern for employees to relate negatively with health/safety outcomes (RQ3).

## Meta-Analytic Procedures

A recent meta-analysis on organizational climate–work group performance relationships (Wallace et al. 2016) provided the foundation for demonstrating the proposed procedures. The criteria for inclusion of primary studies were consistent with the approach of Wallace et al. such that to be included in the current meta-analyses, “we retained studies that (a) reported an effect size between one or more aggregated work climate variables and one or more

performance indicators; (b) presented relationships for climate and criterion variables at the group, team, or unit levels; and (c) provided appropriate justification for aggregation of variables (or enough data to ascertain aggregation suitability)” (Wallace et al. 2016, pp. 847–848). With that said, not all studies in the Wallace et al. (2016) meta-analysis were included herein. For example, Wallace et al. (2016) examined relationships between multiple organizational climate variables and multiple criteria (i.e., worker attitudes, customer satisfaction). Although the Wallace et al. (2016) framework consisted of sorting climate into the higher-order factors of concern for customers and concern for employees, only the higher-order factor of concern for employees was considered for the purposes of the illustrative examples in this study. Also, because the purpose of the current study was to demonstrate a statistical procedure and for reasons of parsimony, only studies that reported relationships between concern for employees and performance were used.

## Coding of Studies

Two of the authors independently coded all studies. Of the 62 samples, 16 did not fit the inclusion criteria. The reasons for excluding studies were that data were only reported at the individual level or the criterion was something other than performance (i.e., attitudes, customer satisfaction, etc.). In cases of disagreement, the coders arrived at consensus through an in-person discussion.

Group-level reliabilities (i.e., ICC(2)s and stability coefficients) were recorded when reported in primary studies. If only an ICC(1) was reported for the predictor, criterion, or both, it was converted into an ICC(2) using the Spearman–Brown formula to estimate the reliability of the group means (Bliese 2000). In the few instances where primary studies reported multiple climate or performance information, a composite correlation was computed using formulas presented in Hunter and Schmidt (2004). Arguably, computing correlations based on composites is more construct valid than correlations based on a single measure (Thoreson et al. 2003).

Organizational climate was coded such that the lower-level factors of concern for employees were coded when available. In addition, we coded for other climate variables that constituted more specific concerns for employees such as bullying, justice, safety, and support, to name a few.

As previously discussed, the criterion domain was separated into three dimensions: behavioral-oriented performance, productivity-oriented outcomes, and health/safety-oriented outcomes. The behavioral performance dimension consisted of any ratings of overall performance, contextual performance, counterproductive performance/deviance, as well as service and task performance. The productivity

<sup>2</sup> The RBNL meta-analytic procedures for correcting individual correlations depart from those proposed by Hunter and Schmidt (2004) by accounting not only for sampling error due to  $N$ , but additionally take into account sampling error in the reliability estimates. As discussed in Raju et al. (1991), Hunter and Schmidt’s (2004) meta-analytic procedures for individually corrected correlations deal with the special case in which reliabilities of predictor and criterion are assumed to be fixed. The reader is referred to Raju et al. (1991) for a detailed description concerning the advantages of employing sample-based meta-analytic procedures over alternative methods.

performance dimension consisted of studies that reported financial (i.e., ratio, sales, etc.) or productivity outcomes (i.e., piece-rate, counts, etc.). Finally, the health/safety dimension of performance included measures of accidents and injuries.

### *Meta-Analytic Calculations*

Following the RBNL procedures for tests of each research question, the mean and variance of corrected correlations were estimated as well as the random effects standard error for the mean corrected correlation (see Burke and Landis 2003). Given the focus in the current study on correcting individual correlations with available sample-based reliabilities, we also computed the individually corrected correlations and their sampling variances. The RBNL meta-analytic procedures were optimal for this study, as reliability information was occasionally missing for the predictor, criterion, or both, in studies examining group-level phenomena. Again, corrections for range restriction were not considered in light of the overall goals of the current investigation.<sup>3</sup>

For analyses with sample-based reliabilities, meta-analyses were conducted with alternative means for estimating missing predictor and criterion reliabilities. In Scenario 1 the average of the available ICC values was used for missing predictor reliabilities and the average of the available criterion reliabilities was used for missing criterion reliability values. These values were .79 for organizational climate, .76 for behavioral-oriented performance, and .84 for both health/safety and productivity outcomes. In sum, Scenario 1 is a situation where averages are used in place of missing predictor and criterion reliabilities and basic sampling error due to sample size ( $N$ ) along with sampling error in the available sample-based predictor and criterion reliabilities is taken into account.

In the Scenario 2, we relied on assumed population reliabilities values for missing predictor and criterion reliabilities. The predictor reliability value was .80 for organizational climate. This value was the average group-level (test–retest) reliability for work perception measures reported in the large-scale studies by Harter et al. (2002, 2010). An assumed population reliability value of .52 was used for behavioral-oriented performance. This value was, in part, based on the average business-unit-level reliability for customer perceptions (which include perceptions of sales personnel performance) reported in the Harter et al. (2002) of .528, and the common assumed population reliability value of .52 for job performance

measures (see LeBreton et al. 2014). The assumed population reliability of .85 for both health/safety and productivity outcomes was based on the average reliabilities for business-unit outcomes in the Harter et al. studies, which ranged from .78 to .93. That is, .85 is approximately at the mid-point of this range in average reliabilities for business-unit outcomes (i.e., productivity outcomes, financial performance measures, etc.). In sum, Scenario 2 is a situation where assumed population reliabilities are used in place of missing predictor and criterion reliabilities and basic sampling error due to sample size ( $N$ ) along with sampling error in the available sample-based predictor and criterion reliabilities is taken into account.

Finally, analyses in Scenario 3 were conducted with only assumed (literature-based) predictor and criterion reliability values for all primary studies. That is, all reliabilities were assumed, fixed population values and set to .80, .52, and .85 for organizational climate, behavior-oriented performance, and health/safety/productivity outcomes, respectively. Scenario 3 is a situation where only sampling due to sample size ( $N$ ) is taken into account.

These three scenarios were created to cover typical situations for how reliability estimates are currently handled in meta-analyses. Together, the alternative analyses allow for comparisons of how individually corrected correlations and their sampling variances are affected by the use of available ICCs and other types of reliabilities versus assumed population reliabilities as well as how meta-analytic results are affected by the use of alternative means for estimating reliabilities.

### *Supplemental Meta-Analyses*

In addition to presenting meta-analytic results from scenarios reflecting natural nuances of group-level data, supplemental meta-analyses were conducted that restricted the reliability data that entered the meta-analysis. These supplemental analyses were intended to further demonstrate how the degree of available reliability information affects parameter estimates. While many possible restrictions could be imposed on the sample reliability data, we chose a 50 % reduction in available reliability information from what was originally reported. Given that supplemental analyses were employed to illustrate how findings can change based on a restricted condition, the retained reliabilities were from the lower portion of the reliability distribution. This restriction was only imposed on the predictor and criterion reliability distributions for studies that provided data for Research Question 1, where there was sufficient reliability data within the primary studies to make a comparison between findings with all sample-based reliability data and a condition with restricted reliability data. This reduction in reliabilities for RQ 1 resulted in a

<sup>3</sup> One study was excluded from the analyses due to a large sample size and the direction of the effect. Results of these analyses with the study included are available upon request from the author.

condition with roughly 40 and 15 % available predictor and criterion reliability information across all primary studies, respectively. These restricted reliability percentages are somewhat reflective of actual percentages of available reliability information for the relationships examined in RQ2 and RQ3.

The reductions in available reliabilities for the supplemental analyses resulted in adjustments to the sample-weighted reliability averages used when a study's predictor and/or criterion reliability value was missing. For Scenario 1, where the average sample-based reliability was used for a missing reliability value, the new values were .58 and .55 for the organizational climate and behavioral performance, respectively. For Scenario 2 where assumed population values were used for missing reliabilities, the new assumed reliability value for organizational climate was .60. This value was chosen as it closely approximates the sample-weighted reliability from Scenario 1 (for restricted reliabilities) and it is reflective of group-level test–retest reliabilities for attitudinal measures reported in the literature (e.g., see Harter et al. 2002). Consistent with arguments made above, the criterion reliability value remained at .52 for behavioral performance, as it is both a literature-based value and one that is consistent with the sample-weighted reliability from Scenario 1 (for restricted reliabilities). For Scenario 3 where all reliabilities are assumed population values, .60 and .52 were employed for all predictor and criterion reliability values, respectively.

### Primary and Meta-Analytic Results

Because the focal climate–performance research questions reflect situations wherein varying proportions of reliability information are reported, results are presented separately for each question. More specifically, a description of the treatment of reliability estimates (assumed fixed vs. sample-based) across the three scenarios and their resulting primary and meta-analytic outcomes will be presented for the test of each relationship. Organizing the results in this fashion is intended to clarify the type of reliability information reported in the primary studies for each meta-analysis and illustrates how reliability was treated when estimating a particular organizational climate–work group performance relationship. For each research question, we highlight how individually corrected correlations and their sampling variances differ (or by contrast, don't change) for primary studies for each of the three scenarios. Subsequent to the presentation of primary results for each research question, we discuss the meta-analytic finding for that organizational climate–performance relationship.

### Research Question 1

RQ1 focused on the relationship between concern for employees and behavioral-oriented performance. Of note, this relationship had the largest number of effect sizes ( $k = 37$ ) relative to tests of the other expected relationships. Following the proposed correction procedures, Eq. 3 was employed when sample-based reliabilities were available for both the predictor and criterion. For Scenario 1, if only predictor reliability information ( $r_{xx}$ ) was missing, then a sample-size-weighted reliability estimate was used assuming a fixed value (i.e., Eq. 6). Similarly, if only criterion reliability ( $r_{yy}$ ) information was missing, then a sample-size-weighted reliability estimate was used (i.e., Eq. 5). If both predictor and criterion information were missing, the sample-size-weighted reliability estimates for both variables were assumed as fixed (i.e., Eq. 7). For Scenario 1, individually corrected correlations and their sampling variances for primary studies included in testing RQ1 are presented in Table 1. These statistics are also presented in Table 1 for Scenario 2 (use of assumed, literature-based reliabilities for missing reliability values) and Scenario 3 (use of assumed population reliabilities for all reliabilities).

Table 1 indicates that the evaluation of RQ1 involves a situation with almost complete sample-based predictor reliabilities (81 %) and partially available criterion reliabilities (30 %). As shown in Table 1, the alternative means for estimating group-level reliabilities had a meaningful impact on the magnitudes of many corrected correlations and their sampling variances. These findings are important as they suggest that researchers could arrive at different conclusions about the relationship between organizational climate and workgroup performance. For instance, when confidence intervals are computed for each of the three scenarios for *Primary Study # 2*, these intervals do not overlap. Specifically, the confidence interval resulting from Scenario 1 was (.74, .86), whereas the confidence intervals for Scenarios 2 and 3 were (.87, 1.0). When these differences compile across all primary studies, conclusions at the meta-analytic level are likely affected.

Associated meta-analytic results for the three scenarios are presented in Table 2. As shown in Table 2, the relationship between concern for employees and behavioral-oriented performance was both positive for each scenario (i.e., .44, .48, and .51 for Scenarios 1, 2, and 3, respectively) and statistically significant, with all confidence intervals excluding zero. Notably, not only did the magnitudes of the relationship between climate and performance differ depending on how reliabilities were estimated, but the confidence intervals also varied across the three scenarios. Use of sample-based reliabilities

**Table 1** Primary study results related to Research Question 1: the relationship between concern for employees and behavioral-oriented outcomes

Primary study	$N_g$	$r_{xx}$	$r_{yy}$	$r$	Scenario 1 $\hat{\rho}$ (sampling variance)	Scenario 2 $\hat{\rho}$ (sampling variance)	Scenario 3 $\hat{\rho}$ (sampling variance)
1	204	.62	.59	.27	.45 (0.010)	.45 (0.010)	.42 (0.010)
2	463	–	–	.62	.80 (0.001)	.96 (0.002)	.96 (0.002)
3	31	.68	–	.54	.75 (0.026)	.91 (0.038)	.84 (0.040)
4	46	–	–	.39	.50 (0.027)	.60 (0.038)	.60 (0.038)
5	56	.95	–	.3	.35 (0.021)	.43 (0.030)	.47 (0.036)
6	26	.81	–	.26	.33 (0.056)	.40 (0.081)	.40 (0.084)
7	160	.51	.98	.25	.35 (0.010)	.35 (0.010)	.39 (0.013)
8	185	.64	.64	.45	.70 (0.006)	.70 (0.006)	.70 (0.008)
9	47	–	–	.45	.58 (0.023)	.70 (0.033)	.70 (0.033)
10	134	.88	–	.39	.48 (0.008)	.58 (0.012)	.60 (0.013)
11	1407	.78	.86	.22	.27 (0.001)	.27 (0.001)	.34 (0.002)
12	67	.99	–	.57	.66 (0.009)	.80 (0.013)	.88 (0.016)
13	150	.97	–	.32	.37 (0.007)	.45 (0.011)	.50 (0.013)
14	81	.47	.27	.27	.76 (0.061)	.76 (0.061)	.42 (0.023)
15	21	.57	–	.66	1.00 (0.015)	1.2 (0.022)	1.0 (0.038)
16	55	.51	–	.04	.06 (0.048)	.08 (0.070)	.06 (0.044)
17	91	.96	–	.28	.33 (0.013)	.40 (0.020)	.43 (0.023)
18	25	.56	.56	.59	1.1 (0.020)	1.1 (0.020)	.91 (0.043)
19	20	.61	–	.54	.79 (0.042)	.96 (0.061)	.84 (0.063)
20	25	.65	–	.49	.70 (0.040)	.84 (0.060)	.76 (0.060)
21	620	.83	.81	.29	.35 (0.002)	.35 (0.002)	.45 (0.003)
22	300	.64	–	.29	.42 (0.005)	.42 (0.005)	.45 (0.007)
23	34	–	–	.3	.39 (0.042)	.47 (0.060)	.47 (0.060)



**Table 1** continued

Primary study	$N_g$	$r_{xx}$	$r_{yy}$	$r$	Scenario 1 $\hat{\rho}$ (sampling variance)	Scenario 2 $\hat{\rho}$ (sampling variance)	Scenario 3 $\hat{\rho}$ (sampling variance)
24	33	.58	–	.64	.96 (0.012)	1.20 (0.020)	.99 (0.026)
25	187	–	–	.24	.31 (0.008)	.37 (0.011)	.37 (0.011)
26	53	–	–	.75	.97 (0.006)	1.2 (0.009)	1.2 (0.009)
27	167	.83	–	.37	.47 (0.007)	.56 (0.010)	.57 (0.011)
28	114	.83	.83	.15	.18 (0.012)	.18 (0.012)	.23 (0.020)
29	160	.71	–	.27	.37 (0.009)	.44 (0.014)	.42 (0.013)
30	56	.94	–	.29	.34 (0.021)	.41 (0.031)	.45 (0.037)
31	134	.47	.47	.26	.55 (0.025)	.55 (0.025)	.40 (0.016)
32	86	.81	.84	.66	.80 (0.004)	.80 (0.004)	1.0 (0.010)
33	48	.95	–	.33	.39 (0.023)	.47 (0.034)	.51 (0.041)
34	111	.76	.58	.36	.54 (0.013)	.54 (0.013)	.56 (0.017)
35	129	.43	–	.06	.10 (0.024)	.13 (0.035)	.10 (0.020)
36	47	–	–	.65	.84 (0.012)	1.0 (0.017)	1.0 (0.017)
37	401	.8	–	.38	.49 (0.003)	.59 (0.004)	.59 (0.004)

Dashes under the predictor and criterion reliability columns indicate a reliability estimate was not reported in the primary study.  $N_g$  is the number of groups or business units.  $r_{xx}$  is the sample-based predictor reliability.  $r_{yy}$  is the sample-based criterion reliability.  $\hat{\rho}$  is the estimate of rho, the individually corrected correlation. Sampling variance refers to the estimated sampling variance of an individually corrected correlation. Scenario 1 refers to the use of sample-based reliabilities and the average sample-based reliability for a missing reliability value. Scenario 2 refers to the use of sample-based reliabilities and an assumed population reliability for a missing reliability value. Scenario 3 refers to the use of only assumed (literature-based) reliabilities, where all reliability values are treated as assumed fixed

resulted in more conservative findings about the relationship.

**Research Question 2**

RQ2 focused on the relationship between concern for employees and productivity-oriented outcomes. This relationship had fourteen reported effects. Following the same procedures used to evaluate RQ1 for Scenario 1, the sample-size-weighted reliability indices were inserted into sampling variance equations wherever reliability information was missing for the predictor, criterion, or both. Again,

if predictor and criterion reliability values were available, then Eq. 3 was employed for estimating the sampling variance of the corrected correlation. If only reliability information was missing for the predictor, then we treated the predictor as assumed fixed (i.e., Eq. 6). If only criterion reliability information was missing, then we treated the criterion as assumed fixed (i.e., Eq. 5). In the situation where both predictor and criterion information were missing, these values were treated as assumed fixed (i.e., Eq. 7). For Scenario 1, individually corrected correlations and their sampling variances associated with RQ2 are presented in Table 3. These statistics are also presented in

**Table 2** Meta-analytic results for tests of relationships between concern for employees and criterion variables

Criterion variable (Research Question)	Treatment of reliabilities	$M_\rho$	$SE_{M_\rho}$	95 % CI		$SD_\rho$
				L	U	
Behavioral-oriented performance (R1)	Scenario 1 (All)	.44	.033	.37	.50	.178
	Scenario 2 (All)	.48	.040	.40	.56	.220
	Scenario 3 (All)	.51	.037	.44	.58	.198
	Scenario 1 (Restricted)	.58	.042	.50	.66	.226
	Scenario 2 (Restricted)	.58	.042	.50	.66	.229
	Scenario 3 (Restricted)	.59	.042	.50	.67	.228
Productivity-oriented outcomes (R2)	Scenario 1	.24	.047	.14	.33	.159
	Scenario 2	.23	.047	.14	.33	.158
	Scenario 3	.23	.043	.14	.31	.146
Health/safety outcomes (R3)	Scenario 1	-.24	.130	-.50	.01	.339
	Scenario 2	-.24	.130	-.50	.01	.340
	Scenario 3	-.23	.114	-.45	-.01	.300

$k$  is the number of independent effect sizes included in each analysis and was 37, 14, and 8 for R1, R2, and R3, respectively.  $N_g$  is the number of groups and was 5974, 3585, and 495 for R1, R2, and R3, respectively.  $M_r$  is the mean observed correlation and was .33, .19, and  $-.19$  for R1, R2, and R3, respectively.  $SD_r$  is the standard deviation of the observed correlations and was .14, .13, and .27 for R1, R2, and R3, respectively.  $M_\rho$  is the mean correlation corrected for unreliability;  $SE_{M_\rho}$  is the standard error of  $M_\rho$ ;  $SD_\rho$  is the standard deviation of corrected correlations; 95 % CI is the 95 % CI for  $M_\rho$ . Scenario 1 refers to the use of sample-based reliabilities and the average sample-based reliability for a missing reliability value. Scenario 2 refers to the use of sample-based reliabilities and an assumed population reliability for a missing reliability value. Scenario 3 refers to the use of only assumed (literature-based) reliabilities, where all reliability values are treated as assumed fixed. These values were .80 and .52 for the predictor and criterion, respectively. For restricted analyses, Scenario 1 (Restricted) refers to an artificially restricted condition wherein there is a 50 % reduction in available reliability and new sample-based average reliabilities are used when missing. These values are .58 and .55 for the predictor and criterion, respectively. Scenario 2 (Restricted) refers to analyses with a 50 % reduction in available reliability and new values are used for the predictor and criterion when missing. These values are .6 and .52 for the predictor and criterion, respectively. Scenario 3 (Restricted) refers to the use of only assumed (literature-based) reliabilities, where all reliability values are treated as assumed fixed. The values under this restricted condition are .6 and .52 for the predictor and criterion, respectively

Table 3 for Scenario 2 (use of assumed reliabilities for missing reliability values) and Scenario 3 (use of assumed population reliabilities for all reliabilities).

Table 3 illustrates a situation with a relatively large proportion (i.e., 79 %) of sample-based predictor reliabilities and a modest percentage of criterion reliabilities (i.e., 36 %). Consistent with findings from RQ1, the alternative approaches for estimating group-level reliabilities had a meaningful impact on the magnitudes of many corrected correlations and their sampling variances.

As shown in Table 2, the relationship between concern for employees and productivity-oriented outcomes was both positive and statistically significant across the three scenarios. These results suggest that greater concern for employees is related to greater levels of workgroup or organizational productivity. Notably, the findings varied little between the scenarios with respect to RQ2 reflecting the fact that the average sample-based reliabilities were comparable in magnitude to the literature-based, assumed population reliabilities for both the predictor and criterion.

### Research Question 3

RQ3 focused on the relationship between concern for employees and health/safety-oriented outcomes. Only eight effects were available for this analysis. As in the previous analyses for Scenario 1, the sample-size-weighted reliability was inserted into sampling variance equations whenever this information was missing. Likewise for Scenario 1, the sample-size-weighted reliability for health/safety outcomes was inserted into sampling variance equations when this value was missing from primary studies. For Scenarios 2 and 3, the same treatments of reliability as described for RQ1 and RQ2 were also applied with these analyses.

Individually corrected correlations and their sampling variances are presented in Table 4 for Scenarios 1, 2, and 3. Table 4 illustrates a situation with relatively fewer sample-based predictor (63 %) and criterion reliabilities (13 %) than either of the prior analyses. As with findings for RQ1 and RQ2, the corrected correlations and sampling

**Table 3** Primary study results related to Research Question 2: the relationship between concern for employees and productivity-oriented outcomes

Primary study	$N_g$	$r_{xx}$	$r_{yy}$	$r$	Scenario 1 $\hat{\rho}$ (sampling variance)	Scenario 2 $\hat{\rho}$ (sampling variance)	Scenario 3 $\hat{\rho}$ (sampling variance)
1	47	–	.52	.45	.70 (0.026)	.70 (0.025)	.55 (0.020)
2	463	–	–	.04	.05 (0.003)	.05 (0.003)	.05 (0.003)
3	136	.72	.93	.38	.46 (0.007)	.46 (0.007)	.46 (0.008)
4	26	.81	–	.24	.29 (0.052)	.30 (0.051)	.30 (0.052)
5	1407	.78	.88	.25	.30 (0.001)	.30 (0.001)	.30 (0.001)
6	150	.97	–	.13	.14 (0.008)	.14 (0.008)	.16 (0.010)
7	121	.48	.87	.38	.59 (0.012)	.59 (0.012)	.46 (0.010)
8	654	.77	–	.25	.31 (0.002)	.31 (0.002)	.30 (0.002)
9	34	–	–	–.25	–.31 (0.040)	–.30 (0.040)	–.30 (0.040)
10	56	.94	–	–.21	–.24 (0.021)	–.23 (0.021)	–.25 (0.024)
11	86	.81	–	–.09	–.11 (0.017)	–.11 (0.020)	–.11 (0.020)
12	72	.72	–	.05	.07 (0.023)	.06 (0.023)	.06 (0.020)
13	129	.43	–	.16	.27 (0.020)	.26 (0.020)	.20 (0.011)
14	204	.62	.51	–.03	–.05 (0.016)	–.05 (0.016)	–.04 (0.007)

Dashes under the predictor and criterion reliability columns indicate a reliability estimate was not reported in the primary study.  $N_g$  is the number of groups or business units.  $r_{xx}$  is the sample-based predictor reliability.  $r_{yy}$  is the sample-based criterion reliability.  $\hat{\rho}$  is the estimate of rho, the individually corrected correlation. Sampling variance is the estimated sampling variance of an individually corrected correlation. Scenario 1 refers to the use sample-based reliabilities and the average sample-based reliability for a missing reliability value. Scenario 2 refers to the use of sample-based reliabilities and an assumed population reliability for a missing reliability value. Scenario 3 refers to the use of only assumed (literature-based) reliabilities, where all reliability values are treated as assumed fixed

variances differed according to how reliability information was handled across the three scenarios.

The meta-analytic results reported in Table 2 regarding RQ3 indicate that conclusions about statistical significance differ depending on whether reliabilities are sample-based or literature-based. For Scenarios 1 and 2, the confidence intervals for  $M_\rho$  include zero and, thus, indicate a statistically nonsignificant relationship between organizational climate and health/safety outcomes. However, in Scenario 3 when all reliabilities are assumed population reliabilities, the confidence interval for the relationship between organizational climate and health/safety outcomes excludes

zero and becomes statistically significant. Again, we caution that the findings pertaining to RQ3 are based on a small number of studies, where there is likely to be more variability due to second-order sampling of studies.

*Supplemental Meta-Analytic Findings*

Results from the supplemental analyses for Research Question 1 with restricted reliabilities for Scenario 1 to Scenario 3 are presented in Table 2. As expected, the magnitudes of  $M_\rho$  are substantially greater when reliability information is restricted to the lower portion of the

**Table 4** Primary study results for Research Question 3: the relationship between concern for employees and health/safety-oriented outcomes

Primary study	$N_g$	$r_{xx}$	$r_{yy}$	$r$	Scenario 1 $\hat{\rho}$ (sampling variance)	Scenario 2 $\hat{\rho}$ (sampling variance)	Scenario 3 $\hat{\rho}$ (sampling variance)
1	81	.47	–	–.35	–.56 (0.021)	–.55 (0.020)	–.42 (0.014)
2	21	.57	–	–.61	–.89 (0.023)	–.88 (0.022)	–.74 (0.030)
3	46	.72	.69	.31	.44 (0.033)	.44 (0.033)	.38 (0.030)
4	15	–	–	.45	.56 (0.069)	.55 (0.070)	.55 (0.070)
5	21	–	–	–.26	–.32 (0.066)	–.32 (0.064)	–.32 (0.064)
6	33	.58	–	.31	.45 (0.049)	.44 (0.050)	.38 (0.040)
7	25	–	–	–.42	–.52 (0.043)	–.51 (0.042)	–.51 (0.042)
8	253	.88	–	–.27	–.32 0.005	–.31 (0.005)	–.33 (0.005)

Dashes under the predictor and criterion reliability columns indicate a reliability estimate was not reported in the primary study.  $N_g$  is the number of groups or business units.  $r_{xx}$  is the sample-based predictor reliability.  $r_{yy}$  is the sample-based criterion reliability.  $\hat{\rho}$  is the estimate of rho, the individually corrected correlation. Sampling variance is the estimated sampling variance of an individually corrected correlation. Scenario 1 refers to the use sample-based reliabilities and the average sample-based reliability for a missing reliability value. Scenario 2 refers to the use of sample-based reliabilities and an assumed population reliability for a missing reliability value. Scenario 3 refers to the use of only assumed (literature-based) reliabilities, where all reliability values are treated as assumed fixed

reliability distribution. Moreover, for Scenario 1, the reduction in available reliability information greatly affects the standard error of  $M_\rho$  (i.e., the square root of the sampling variance of  $M_\rho$ ), increasing it by 33 % from the original estimate of  $SE_{M_\rho}$ . As noted above, there are 50 % fewer original predictor and criterion reliability sampling variances included in the estimate of the  $SE_{M_\rho}$  with restricted reliability data. Notably, this reduction in sample-based reliabilities and their respective sampling variances results in not only the 33 % increase in  $SE_{M_\rho}$ , but it also leads to a 95 % confidence interval for  $M_\rho$  that very minimally overlaps with the original 95 % confidence interval for  $M_\rho$ . Further reductions in sample-based reliabilities and their sampling variances, not reported here, produce even more marked changes in the original and revised estimates of  $SE_{M_\rho}$  and substantive conclusions about the relationship between organizational climate and behavioral performance at the group-level of analysis.<sup>4</sup>

<sup>4</sup> Supplemental meta-analyses were also conducted with a 75 % reduction in available reliability from the original amount and are available upon request from the first author. Most notably, when sample-based reliabilities (and their associated sampling variances) are artificially extracted we see a marked increase in  $SE_{M_\rho}$  and a substantially larger magnitude of  $M_\rho$ . Furthermore, the range of

For Scenarios 2 and 3 with restricted reliability data, the results are very similar to those for Scenario 1 with restricted reliability data. These findings are not surprising given that the primary difference between these scenarios was somewhat lower assumed population predictor reliability for missing reliabilities in Scenarios 2 and all predictor reliabilities in Scenario 3 (i.e., .6 vs. .8). However, a comparison of the original estimates of  $M_\rho$  and  $SE_{M_\rho}$  and the respective estimates based on restricted reliability data for Scenarios 2 and 3 indicates meaningful differences. Together, these differences result in original and revised confidence intervals for  $M_\rho$  that are considerably divergent.

## Discussion

This paper presented arguments for treating ICC(2) information in primary studies as an estimate of group-level reliability when correcting correlations for unreliability in measures of group- or business-unit-level phenomena, and employing sampling variance equations presented by Raju

Footnote 4 continued  
confidence moves farther away from original surrounding calculations when the data naturally occurred.

and colleagues (Raju and Brand 2003; Raju et al. 1991) to account for sampling error in group-level reliability estimates. Importantly, our illustrative analyses indicated how findings pertaining to both individually corrected correlations and meta-analytic results can change depending on assumptions about and usage of available ICCs versus assumed population reliability. These differences were evident in the magnitudes of corrected correlations and sampling variances in both primary studies and meta-analyses. Notably, the use of available ICCs and other types of group-level reliabilities tended to produce more conservative estimates of relationships between variables within primary studies. Although the differences in the magnitudes of findings were not as pronounced with respect to the meta-analyses, it is important to note that they could lead a researcher to draw different conclusions. For instance, conclusions about the strength of the relationship between organizational climate and behaviorally oriented performance and the statistical significance of the relationship between organizational climate and health/safety outcomes differed depending on the use of sample-based versus literature-based (assumed population) reliabilities.

The point that conclusions about substantive relationships between variables can change based on availability of sample-based reliabilities was further illustrated in supplemental meta-analyses with restricted reliability data. Notably, a 50 % reduction in sample-based reliabilities produced an estimated organizational climate–behavioral performance relationship that not only differed considerably in magnitude from the original estimate, but also one where there was very minimal overlap in confidence intervals for  $M_\rho$ . Importantly, these differences resulted solely from reducing the proportion of sample-based reliabilities from the original to the supplemental analyses. The differences between the original and supplemental analyses more conclusively illustrate how the availability and treatment of group-level reliabilities impact meta-analytic findings.

We note that while our paper is not directed toward the specific assumed reliability values to employ in the absence of sample-based group-level reliabilities, our demonstrations indicate the importance of using more conservative assumed reliability values in situations where either sample-based reliabilities are not reported or missing altogether. In particular, for RQ1, the findings across the three scenarios suggest that researchers should rely on the average sample-based reliability (as opposed to an assumed literature-based reliability) for missing reliability values when meaningful percentages (e.g., 30 % or more) of the sample-based predictor and criterion reliabilities are available. That is, the findings from Scenario 1, which relied on the average sample-based

reliability value for missing reliabilities, were more conservative in comparison with the findings for Scenarios 2 and 3 that relied on assumed population values for missing reliabilities.

Importantly, our illustrative analyses provide guidance to primary and meta-analytic researchers in regard to how to correct group-level correlations for unreliability in the predictor, criterion, or both whenever and in whatever proportions the artifact information is available. As such, the conduct of these demonstrations is consistent with LeBreton et al. (2014) and Burke et al.'s (2014) calls for using more accurate and reasonable reliability estimates when correcting correlations for measurement error. Also, our work suggests a need for primary researchers to attend more to estimating and reporting the reliability of aggregated measures to facilitate the estimation of corrected correlations. This need is particularly evident in regard to the reporting of criterion reliabilities in group-level studies, as the largest percentage of available criterion reliabilities for any meta-analysis included herein was 36 %. As noted above, in scenarios where predictor and criterion reliability information is missing, our findings point to the need to give consideration to the use of available (average) ICCs and other sample-based reliabilities as opposed to the reliance on assumed, literature-based population reliabilities. While our findings do not address the accuracy of parameter estimates, they do point to the possible conservative nature of findings based on the use of sample-based reliabilities in both primary and meta-analytic studies.

In conclusion, this investigation presented a rationale and procedures for using sample-based predictor and criterion reliability information to estimate relationships between group-level variables within primary and meta-analytic studies. Given that lack of complete reliability data is a common problem in psychological and organizational research, a primary contribution of this study is the illustration of how to handle situations with varying levels of reported group reliability information and the implications of assumptions made (i.e., treatment of reliabilities as assumed fixed vs. sample-based) when correcting correlations based on aggregated data. Our understanding regarding the accuracy of meta-analytic findings at the group-level can be advanced through future simulation work examining different levels of availability of reliability information, different means for estimating population-level reliabilities for primary studies with missing reliability values, and with respect to the number of respondents per group. These recommendations are offered in the spirit of improving meta-analytic parameter estimates and our understanding of relationships between constructs at the group-level of analysis.

## References

- \*Studies included in meta-analytic calculations
- Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1992. *Journal of Applied Psychology, 77*, 836–874. doi:10.1037/0021-9010.77.6.836.
- \*Baer, M., & Frese, M. (2003). Innovation is not enough: Climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior, 24*, 45–68. doi:10.1002/job.179.
- Bartko, J. J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin, 83*, 762–765. doi:10.1037/0033-2909.83.5.762.
- Bliese, P. D. (2000). Within group-agreement, non-independence, and reliability: Implications for data aggregation and analysis. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). San Francisco, CA: Jossey-Bass.
- \*Borucki, C. C., & Burke, M. J. (1999). An examination of service-related antecedents to retail store performance. *Journal of Organizational Behavior, 20*, 943–962. doi:10.1002/(SICI)1099-1379(199911)20:6<943::AID-JOB976>3.0.CO;2-9.
- Burke, M. J., & Landis, R. S. (2003). Methodological and conceptual challenges in conducting and interpreting meta-analyses. In K. Murphy (Ed.), *Validity generalization: A critical review* (pp. 287–309). Mahwah, NJ: Erlbaum.
- Burke, M. J., Landis, R. J., & Burke, M. I. (2014). 80 and beyond: Recommendations for disattenuating correlations. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 7*, 531–535.
- \*Chen, G., Kirkman, B., Kanfer, R., Allen, D., & Rosen, B. (2007). A multilevel study of leadership, empowerment, and performance in teams. *Journal of Applied Psychology, 92*, 331–346. doi:10.1037/0021-9010.92.2.331.
- Christian, M. S., Bradley, J. C., Wallace, J., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology, 94*(5), 1103–1127. doi:10.1037/a0016172.
- \*Collins, C. J., & Smith, K. G. (2006). Knowledge exchange and combination: The role of human resource practices in the performance of high-technology firms. *Academy of Management Journal, 49*, 544–560. doi:10.5465/AMJ.2006.21794671.
- \*Colquitt, J. A., Noe, R. A., & Jackson, C. L. (2002). Justice in teams: Antecedents and consequences of procedural justice climate. *Personnel Psychology, 55*, 83–109. doi:10.1111/j.1744-6570.2002.tb00104.x.
- Cronbach, L. J., Gleser, G., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York, NY: Wiley.
- \*Dawson, J. F., Gonzalez-Roma, V., Davis, A., & West, M. A. (2008). Organisational climate and climate strength in UK hospitals. *European Journal of Work and Organizational Psychology, 17*, 89–111. doi:10.1080/13594320601046664.
- \*de Jong, A., de Ruyter, K., & Lemmink, J. (2005). Service climate in self-managing teams: Mapping the linkage of team member perceptions and service performance outcomes in a business-to-business setting. *Journal of Management Studies, 42*, 1593–1620. doi:10.1111/j.1467-6486.2005.00558.x.
- \*Dietz, J., Pugh, S. D., & Wiley, J. (2004). Service climate effects on customer attitudes: An examination of boundary conditions. *Academy of Management Journal, 47*, 81–92. doi:10.2307/20159561.
- Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika, 16*, 407–424. doi:10.1007/BF02288803.
- \*Ehrhart, M. G. (2004). Leadership and procedural justice climate as antecedents of unit-level organizational citizenship behavior. *Personnel Psychology, 57*, 61–94. doi:10.1111/j.1744-6570.2004.tb02484.x.
- \*Fay, D., Luhrmann, H., & Kohl, C. (2004). Proactive climate in a post-reorganization setting: When staff compensate managers' weakness. *European Journal of Work and Organizational Psychology, 13*, 241–267. doi:10.1080/13594320444000083.
- \*Gelade, G. A., & Ivery, M. (2003). The impact of human resource management and work climate on organizational performance. *Personnel Psychology, 56*, 383–404. doi:10.1111/j.1744-6570.2003.tb00155.x.
- \*Gelade, G. A., & Young, S. (2005). Test of a service profit chain model in the retail banking sector. *Journal of Occupational & Organizational Psychology, 78*, 1–22. doi:10.1348/096317904X22926.
- \*Gil, F., Rico, R., Alcover, C. M., & Barrasa, A. (2005). Change-oriented leadership, satisfaction and performance in work groups. *Journal of Managerial Psychology, 20*, 312–328. doi:10.1108/02683940510589073.
- \*Gonzalez-Roma, V., Fortes-Ferreira, L., & Peiro, J. (2009). Team climate, climate strength and team performance. A longitudinal study. *Journal of Occupational and Organizational Psychology, 82*, 511–536. doi:10.1348/096317908X370025.
- Harter, J. K., Schmidt, F. L., Asplund, J. W., Killham, E. A., & Agrawal, S. (2010). Causal impact of employee work perceptions on the bottom line of organizations. *Perspectives on Psychological Science, 5*(4), 378–389. doi:10.1177/1745691610374589.
- Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology, 87*(2), 268–279. doi:10.1037/0021-9010.87.2.268.
- \*Hofmann, D. A., & Mark, B. (2006). An investigation of the relationship between safety climate and medication errors as well as other nurse and patient outcomes. *Personnel Psychology, 59*, 847–869. doi:10.1111/j.1744-6570.2006.00056.x.
- \*Hofmann, D. A., & Stetzer, A. (1996). A cross-level investigation of factors influencing unsafe behaviors and accidents. *Personnel Psychology, 49*, 307–339. doi:10.1111/j.1744-6570.1996.tb01802.x.
- Hong, Y., Liao, H., Hu, J., & Jiang, K. (2013). Missing link in the service profit chain: A meta-analytic review of the antecedents, consequences, and moderators of service climate. *Journal of Applied Psychology, 98*(2), 237–267. doi:10.1037/a0031666.
- \*Hui, C. H., Chiu, W. C. K., Yu, P. L. H., Cheng, K., & Tse, H. (2007). The effects of service climate and the effective leadership behaviour of supervisors on frontline employee service quality: A multi-level analysis. *Journal of Occupational & Organizational Psychology, 80*, 151–172. doi:10.1348/096317905X89391.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- \*Katz-Navon, T., Naveh, E., & Stern, Z. (2005). Safety climate in healthcare organizations: A multidimensional approach. *Academy of Management Journal, 48*, 1075–1089. doi:10.5465/AMJ.2005.19573110.
- \*King, E. B., de Chermont, K., West, M., Dawson, J. F., & Hebl, M. (2007). How innovation can alleviate negative consequences of demanding work contexts: The influence of climate for innovation on organizational outcomes. *Journal of Occupational and Organizational Psychology, 80*, 631–645. doi:10.1348/096317906X171145.

- Lahey, M., Downey, R. G., & Saal, F. E. (1983). Intraclass correlations: There's more there than meets the eye. *Psychological Bulletin*, 93(3), 586–595. doi:10.1037/0033-2909.93.3.586.
- LeBreton, J. M., Scherer, K. T., & James, L. R. (2014). Corrections for criterion reliability in validity generalization: A false prophet in a land of suspended judgment. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 74, 478–500.
- \*Liao, H., & Chuang, A. (2004). A multilevel investigation of factors influencing employee service performance and customer outcomes. *Academy of Management Journal*, 47, 41–58. doi:10.2307/20159559.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Menlo Park, CA: Addison-Wesley.
- \*Luria, G. (2008). Controlling for quality: Climate, leadership, and behavior. *Quality Management Journal*, 15, 27–42.
- \*Luria, G., & Yagil, D. (2008). Procedural justice, ethical climate and service outcomes in restaurants. *International Journal of Hospitality Management*, 27, 276–283. doi:10.1016/j.ijhm.2007.07.022.
- \*MacCormick, J. S., & Parker, S. K. (2010). A multiple climates approach to understanding business unit effectiveness. *Human Relations*, 63, 1171–1806. doi:10.1177/0018726710365090.
- \*Mayer, D. M., Kuenzi, M., & Greenbaum, R. L. (2010). Examining the link between ethical leadership and employee misconduct: The mediating role of ethical climate. *Journal of Business Ethics*, 95, 7–16. doi:10.1007/s10551-0110-794-0.
- \*Maynard, M. T., Mathieu, J. E., Marsh, W. M., & Ruddy, T. M. (2007). A multilevel investigation of the influences of employees' resistance to empowerment. *Human Performance*, 20, 147–171. doi:10.1080/08959280701332885.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46. doi:10.1037/1082-989X.1.1.30.
- \*McKay, P. F., Avery, D. R., & Morris, M. A. (2009). A tale of two climates: Diversity climate from subordinates' and managers' perspectives and their role in store unit sales performance. *Personnel Psychology*, 62, 767–791. doi:10.1111/j.1744-6570.2009.01157.x.
- \*Naumann, S., & Bennett, N. (2002). The effects of procedural justice climate on work group performance. *Small Group Research*, 33, 361–377. doi:10.1177/10496402033003004.
- \*Naveh, E., Katz-Navon, T., & Stern, Z. (2005). Treatment errors in healthcare: A safety climate approach. *Management Science*, 51, 948–960. doi:10.1287/mnsc.1050.0372.
- \*Neal, A., & Griffin, M. A. (2006). A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *Journal of Applied Psychology*, 91, 946–953. doi:10.1037/0021-9010.91.4.946.
- \*Ostroff, C., Kinicki, A., & Clark, M. (2002). Substantive and operational issues of response bias across levels of analysis: An example of climate-satisfaction relationships. *Journal of Applied Psychology*, 87, 355–368. doi:10.1037/0021-9010.87.2.355.
- \*Pirola-Merlo, A., Hartel, C., Mann, L., & Hirst, G. (2002). How leaders influence the impact of affective events on team climate and performance in R&D teams. *Leadership Quarterly*, 13, 561–581. doi:10.1016/S1048-9843(02)00144-3.
- Raju, N. S., & Brand, P. A. (2003). Determining the significance of correlations corrected for unreliability and range restriction. *Applied Psychological Measurement*, 27, 52–71. doi:10.1177/0146621602239476.
- Raju, N. S., Burke, M. J., Normand, J., & Langlois, G. M. (1991). A new meta-analytic approach. *Journal of Applied Psychology*, 76, 432–446. doi:10.1037/0021-9010.76.3.432.
- Raju, N. S., Pappas, S., & Williams, C. P. (1989). An empirical Monte Carlo test of the accuracy of the correlation, covariance, and regression slope models for assessing validity generalization. *Journal of Applied Psychology*, 74, 901–911. doi:10.1037/0021-9010.74.6.901.
- \*Richardson, H. A., & Vandenberg, R. J. (2005). Integrating managerial perceptions and transformational leadership into a work-unit level model of employee involvement. *Journal of Organizational Behavior*, 26, 561–589. doi:10.1002/job.329.
- \*Salanova, M., Agut, S., & Peiro, J. M. (2005). Linking organizational resources and work engagement to employee performance and customer loyalty: The mediation of service climate. *Journal of Applied Psychology*, 90, 1217–1227. doi:10.1037/0021-9010.90.6.1217.
- \*Schmit, M. J., & Allscheid, S. P. (1995). Employee attitudes and customer satisfaction: Making theoretical and empirical connections. *Personnel Psychology*, 48, 521–536. doi:10.1111/j.1744-6570.1995.tb01768.x.
- \*Schneider, B., Ehrhart, M. G., Mayer, D. M., Saltz, J. L., & Niles-Jolly, K. (2005). Understanding organization-customer links in service settings. *Academy of Management Journal*, 48, 1017–1032. doi:10.5465/AMJ.2005.19573107.
- \*Schneider, B., White, S. S., & Paul, M. C. (1998). Linking service climate and customer perceptions of service quality: Test of a causal model. *Journal of Applied Psychology*, 83, 150–163. doi:10.1037/0021-9010.83.2.150.
- \*Schulte, M., Ostroff, C., Shmulyian, S., & Kinicki, A. (2009). Organizational climate configurations: Relationships to collective attitudes, customer satisfaction, and financial performance. *Journal of Applied Psychology*, 3, 618–634. doi:10.1037/a0014365.
- \*Seibert, S. E., Silver, S. R., & Randolph, W. A. (2004). Taking empowerment to the next level: A multiple-level model of empowerment, performance, and satisfaction. *Academy of Management Journal*, 47, 332–349. doi:10.2307/20159585.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. doi:10.1037/0033-2909.86.2.420.
- \*Simons, T., & Roberson, Q. (2003). Why managers should care about fairness: The effects of aggregate justice perceptions on organizational outcomes. *Journal of Applied Psychology*, 88, 432–443. doi:10.1037/0021-9010.88.3.432.
- \*Smith, K. G., Collins, C. J., & Clark, K. D. (2005). Existing knowledge, knowledge creation capability, and the rate of new product introduction in high-technology firms. *Academy of Management Journal*, 48, 346–357. doi:10.5465/AMJ.2005.16928421.
- \*Sowinski, D. R., Fortmann, K. A., & Lezotte, D. V. (2008). Climate for service and the moderating effects of climate strength on customer satisfaction, voluntary turnover, and profitability. *European Journal of Work and Organizational Psychology*, 17, 73–88. doi:10.1080/13594320701473065.
- Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.). Washington, DC: American Council on Education.
- \*Stetzer, A., Morgeson, F., & Anderson, E. (1997). Organizational climate and ineffectiveness: Evidence from 25 outdoor work-crew divisions. *Journal of Quality Management*, 2, 251–265. doi:10.1016/S1084-8568(97)90006-7.
- Thoreson, C. J., Kaplan, S. A., Barsky, A. P., Warren, C. R., & de Chermont, K. (2003). The affective underpinnings of job perceptions and attitudes: A meta-analytic review and integration. *Psychological Bulletin*, 129, 914–945. doi:10.1037/0033-2909.129.6.914.
- \*Towler, A., Lezotte, D. V., & Burke, M. J. (2011). An examination of the service climate-firm performance chain: The role of customer retention. *Human Resource Management*, 50, 391–406. doi:10.1002/hrm.20422.

- \*Wagner, S. H., Parker, C. P., & Christiansen, N. D. (2003). Employees that think and act like owners: Effects of ownerships beliefs and behaviors on organizational effectiveness. *Personnel Psychology*, *56*, 847–871. doi:[10.1111/j.1744-6570.2003.tb00242.x](https://doi.org/10.1111/j.1744-6570.2003.tb00242.x).
- Wallace, J. C., Edwards, B. D., Paul, J., Burke, M. J., Christian, M., & Eissa, G. (2016). Change the referent? A meta-analytic investigation of direct and referent-shift consensus models for organizational climate. *Journal of Management*, *42*, 838–861.
- \*Wallace, C., & Chen, G. (2006). A multilevel integration of personality, climate, self-regulation, and performance. *Personnel Psychology*, *59*, 529–557. doi:[10.1111/j.1744-6570.2006.00046.x](https://doi.org/10.1111/j.1744-6570.2006.00046.x).
- Whitman, D. S., Caleo, S., Carpenter, N. C., Horner, M. T., & Bernerth, J. B. (2012). Fairness at the collective level: A meta-analytic examination of the consequences and boundary conditions of organizational justice climate. *Journal of Applied Psychology*, *97*(4), 776–791. doi:[10.1037/a0028021](https://doi.org/10.1037/a0028021).
- Whitman, D. S., Van Rooy, D. L., & Viswesvaran, C. (2010). Satisfaction, citizenship behaviors, and performance in work units: A meta-analysis of collective construct relations. *Personnel Psychology*, *63*(1), 41–81. doi:[10.1111/j.1744-6570.2009.01162.x](https://doi.org/10.1111/j.1744-6570.2009.01162.x).
- \*Zohar, D., & Luria, G. (2005). A multilevel model of safety climate: Cross level relationships between organization and group-level climates. *Journal of Applied Psychology*, *90*, 616–628. doi:[10.1037/0021-9010.90.4.616](https://doi.org/10.1037/0021-9010.90.4.616).