2017

Comparisons of Artifact Correction Procedures for Meta-Analysis: An Empirical Examination on Correcting Reliabilities

Lei Zhao
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

COMPARISONS OF ARTIFACT CORRECTION PROCEDURES FOR META-ANALYSIS: AN EMPIRICAL EXAMINATION ON CORRECTING RELIABILITIES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

PROGRAM IN RESEARCH METHODOLOGY

BY

LEI ZHAO

CHICAGO, ILLINOIS

MAY 2017

ACKNOWLEDGMENTS

Completing this Ph.D. degree is the most challenging yet rewarding experience in my life thus far. I am heartily appreciative and grateful to those who have helped and supported me during this journey. Without them, I would never be able to get to where I am now. First, I would like to express my sincere gratitude to my academic advisor, dissertation chair, and a great friend Dr. Meng-Jia Wu. She has supported me not only academically but also emotionally and spiritually. Her guidance, encouragement and immense knowledge not only made this dissertation possible, but also enabled me to learn to become a better person. Also, I would like to thank my dissertation committee members, Dr. Ken Fujimoto and Dr. Tim O'Brien. They both offered immeasurable amount of wisdom and willingness to support me to make the completion of this dissertation possible, and inspired me with numerous research ideas for future studies!

Special thanks should go to my parents Guiling Zuo and Fuli Zhao, and my sister Xin Zhao. They are so loving and understanding to me. Without their support and sacrifice over these years, I would never have been able to pursue my dreams or get to where I am now. Alongside them, I have to thank all my friends who have been there with me through all my ups and downs in the past few years. Lastly, I would like to dedicate this dissertation to all the great people in my life whom I am so thankful to. I wish you all the best in pursuing your life dreams and happiness.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study reviewed some challenges and issues in artifact correction meta-analysis, particularly around using reliability estimates to correct for measurement error. Two individual correction procedures—the Hunter-Schmidt procedure and the procedure developed by Raju, Burke, Normand, and Langlois (the RBNL procedure)—are addressed in this research. The purpose of this study is to use real-world data to examine the differences between meta-analytic estimations produced by the two artifact correction procedures and those by the traditional bare-bones meta-analysis procedures, under the condition of inter-dependent reliabilities. The impact of this inter-correlation on meta-analysis results needs investigation when artifact indicators, such as reliability of predictor and reliability of outcome, are proven to be significantly inter-correlated. The current study revealed that neither the choice of artifact correction nor the choice of analysis procedure provided any significant differences in the estimation results, whereas it was the choice of the reliability estimates that generated noticeable differences in the results. In addition, the violation of the assumption for independent reliability did not greatly impact the meta-analytic estimation results.

CHAPTER ONE

INTRODUCTION

Background

Within the domain of personnel selection, validity generalization (VG) assesses the generalizability of a correlation (i.e., validity coefficient ρ) between employment test score (X) and job performance (Y) across various organizational settings (Callender & Osburn, 1980; Schmidt & Hunter, 1977). Schmidt, Hunter, and their colleagues suggested that much of the observed variation among validity study results is due to statistical artifacts, including sample error, measurement error, range restriction, etc. These artifacts attenuate the true validity coefficient by a multiplicative fraction (Schmidt, Hunter, & Urry, 1976; Schmidt & Hunter, 1977). Thus, an accurate estimation for population validity coefficient can only be obtained if the artifactual attenuations are eliminated from the observed coefficients of individual studies. Schmidt and Hunter (Hunter & Schmidt, 1990; 2004), therefore, proposed a correlation-based meta-analysis method that corrects the point-estimate of the validity coefficient and its variances for artifactual effects in individual studies.

Early research on correcting artifactual effects in validity estimation assumed the nature of independence among artifacts (Hunter & Schmidt, 2004, p. 75; Raju & Burke, 1983; Raju, Anselmi, Goodman, & Thomas, 1998), specifically meaning that artifacts are independent of each other and independent of the true population correlation. However, this assumption in general can hardly be recognized in practice especially because some of the artifacts are mathematically and functionally related. Previous research has particularly examined the interdependent relationship between reliabilities and range restrictions, as well as the interdependency between sampling errors of observed correlations and other artifacts (Hunter, Schmidt, & Le, 2006; Li, 2013; Mendoza & Mumford, 1987; Oswald & Johnson, 1998, Raju & Burke, 1983; Raju, Normand, Burke, & Langlois, 1991; Raju et al., 1998;). As a result of these examinations, formulae and procedures for artifact correction meta-analysis have gone through continuous development and refinement in order to take account of assorted forms of correlated artifacts.

While the research in artifact correction continues, violation of artifact independence and its impact on meta-analysis results in practical research have not yet gained sufficient attention, particularly for the interdependent relationship between reliability of predictor (*X*) and reliability of criterion (*Y*). Recently, this specific assumption for independent reliabilities was reemphasized and empirically examined for its tenability in practice (Köhler, Cortina, Kurtessis, & Gölz, 2015). Although psychometricians are aware that different types of reliability estimates are neither interchangeable nor equivalent to each other under most practical situations (Desimone, 2014), not much research has been put toward the meta-analysis situation wherein mixed types of reliability estimates are used across individual studies. Especially now knowing it is also problematic that the degree of interdependence among reliability estimates varies across different types of estimates. All in all, there is a lack of thorough discussion and practical investigation around the impact of inter-correlated reliabilities and mixed types of reliability estimates on artifact correction results.

The purpose of this research is to use real-world empirical data to examine the impact of correcting for measurement error on meta-analysis estimates under the condition of inter-correlated reliabilities. While a Monte Carlo study by Raju and his colleagues (Raju et al., 1998) theoretically examined this matter in brief using simulated data, a case on this topic using real-world data has not been fully vetted. Expressly, using Hunter-Schmidt's artifact correction procedure (Hunter & Schmidt, 1990; 2004) and Raju-Normand-Burke-Langlois's procedure (Raju et al., 1991; Raju & Brand, 2003) to examine the data from published and unpublished studies, the current research will not only inspect whether there is an effect of correlated reliabilities on meta-analysis results, but also how this effect is manipulated by the inter-correlated relationship amongst different types of reliability estimates.

Conceptual Underpinnings for the Study

Correction for Measurement Error

In current VG research practice, artifact correction for sampling error and measurement error has been well recognized and popularly applied (Aguinis, Dalton, Bosco, Pierce, & Dalton, 2011; Hunter & Schmidt, 2004; Murphy, 2003). As information for sample sizes and reliability estimates becomes more accessible across social science studies, correcting sampling error and measurement error has correspondingly been redeemed as a well-accepted convention in meta-analysis (Murphy, 2003). As a result, an appropriate use of reliability estimates for correcting for measurement error started attracting more attention in contemporary meta-analysis research (Köhler et al., 2015). Hunter et al. (2006) explicitly addressed the consequences of using inappropriate reliability estimates for VG studies. Yet, limited discussions have continued around the consequences of artifact corrections for measurement error using mixed types of reliability estimates (Hunter & Schmidt, 2004). Moreover, there is a lack of practical implications in meta-analysis for the tenability of the assumption for reliability independence (Köhler et al., 2015). The current study hence sets its focus on drawing more attention to this research topic and fulfilling this gap.

Issues in correlated reliability. The methods for correcting for measurement error in meta-analysis assume reliability independence (Hunter & Schmidt, 2004). Köhler et al. (2015) tested the tenability of this assumption through examining the correlation between reliability estimates of two variables across hundreds of empirical studies. Two of their investigations, using real-world data, showed that there were substantial correlations between reliabilities of the perceived organizational support (POS) and the reliabilities of its outcome variables. Also, the degree of this relationship varied between different types of reliability estimates for independent and dependent variables. This interdependent relationship between reliability estimates of two variables could potentially cause a biased correction for the observed bivariate correlation.

Different types of reliability estimate different sources of measurement error, therefore correction for measurement error in meta-analysis should be carefully implemented. Murphy, specifically, stated that “It does not make sense to correct one correlation coefficient for errors due to instability over time and another correlation coefficient for errors due to disagreements between raters” (Murphy, 2003, p. 391). When certain studies are designed in a way that the source of measurement error cannot possibly affect variances in correlation coefficients, it is actually unnecessary to correct correlations for the artifact under inspection. In summary, if correction for measurement error is carelessly conducted, there will be a risk of removing variances due to situational moderators (Murphy, 2003, p. 391). Surprisingly, there has not been much of practical investigation in VG meta-analysis focusing on correcting for measurement error by taking into consideration of the types of reliability estimates.

Artifact Correction Procedures

In general, there are two artifact correction approaches meta-analysis researchers often employ, individual correction approach and artifact-distribution-based correction approach. The choice of these two approaches mainly depends on how much information is available for statistical artifacts. Individual correction can be applied if data (such as sample size, reliability estimates, range restriction, etc.) is available in each sample study included in a meta-analysis. Otherwise, when most studies do not report or sparsely report information on artifacts, yet one still wishes for correction, then one can rely on the distributions of these artifacts to continue the artifact correction meta-analysis in their study.

The two approaches. The present study sets its focus on the individual correction approach for meta-analysis. The use of artifact-distribution-based correction has mainly received praise for its practicality in handling missing values, as well as for its usefulness in incorporating relevant information from studies that are not part of the current meta-analysis. However, as the American Psychological Association (APA) continues to advocate standardized reporting practice in social science research, information availability has been improved tremendously in study reports, particularly for variables of interest and their measures, sample sizes, and reliability estimates, though not for range restriction. As a result, the observed phenomenon is that the individual correction approach has been more popularly adopted by recent meta-analysis studies rather than the artifact-distribution-based correction approach (Aytug, Rothstein, Zhou, & Kern, 2012). Secondly, comparison studies provided empirical evidences showing that the estimation results of artifact-distribution-based correction are not yet as accurate as those of individual correction approach where each observed correlation coefficient is corrected and the population estimates are derived from the corrected individual correlations (Hunter & Schmidt, 2004, p. 157; Murphy, 2003, p. 304). Thirdly, an inherent disadvantage of distribution-based correction is that it demands an important yet unrealistic assumption of artifact independence across studies, whereas certain individual correction procedures do not necessarily require this assumption.

The RBNL procedure. One of the individual correction procedures developed by Raju, Burke, Normand, and Langlois in 1991 (Raju et al., 1991), termed the RBNL procedure, produces unbiased population estimates for a bivariate correlation of true scores. Unlike other individual correction procedures, the RBNL procedure was developed to specifically take account of sampling errors of the artifacts themselves, with consideration that available artifacts across studies often suffer from sampling errors too. This procedure improves the accuracy of population estimates, especially in contrast to the results from other artifact correction procedures such as the procedure of Taylor Series Approximations 1 (Raju et al., 1991). In addition, the RBNL procedure pragmatically requires fewer assumptions and addresses various scenarios of missing artifacts. The most important thing is that the RBNL procedure was stated as a suitable artifact correction method regardless of the types of reliabilities. “Instead of offering a separate sampling variance formula for each definition of reliability, this development proposes a single formula for the sampling variance of corrected corrections, which may be used with any definition of reliability” (Raju & Brand, 2003). However, the RBNL procedure has not been applied as popularly as the Hunter-Schmidt procedures (Hunter & Schmidt, 1990; 2004) because of the domination of the artifact-distribution-based approach in general, as well as the academic influence of Frank Schmidt and John Hunter the pioneers of artifact correction. Aguinis et al. (2011) reviewed 192 published meta-analysis studies between1982 and 2009 in which a total of more than 5,000 correlation effect sizes were included; however, only about three percent of the effects sizes were synthesized using the RBNL procedure.

The Hunter-Schmidt procedure. In the belief of artifact correction, Schmidt and Hunter specifically dedicated a thorough discussion regarding the use of reliability estimates for correcting for measurement error (Schmidt & Hunter, 1996). They described 26 different research scenarios to address the importance and necessity of correcting for measurement error and the use of appropriate types of reliability estimates for correction. They advocated that researchers should estimate reliability properly when conducting individual studies, in which way these "appropriate" estimates can be used for correcting for measurement error in a meta-analysis study. But they did not explicitly describe how their artifact correction procedures can assure the technical accuracy when various kinds of reliability estimates for *X* and *Y* are already reported in individual studies and ready to be entered into the correction formulae for meta-analysis. In addition, the effect of interdependence between reliability estimates of *X* and *Y* on meta-analytic estimations has rarely been researched or discussed by Hunter, Schmidt, and their followers (Hunter & Schmidt, 1990; 2004). With the increasing popularity of Hunter-Schmidt correction procedures in meta-analysis, it is necessary to test how the estimation results drawn from their procedures are impacted by different degrees of reliability interdependence across various types of reliability estimates.

The main research focus of this study is to use real-world data to compare the RBNL and the Hunter-Schmidt artifact correction procedures, and demonstrate how they perform similarly or differently based on the real-world empirical data, under the condition of different degrees of reliability interdependence across various types of reliability estimates. The analysis results from this study will also be compared to conclusions from some Monte Carlo studies in order to examine whether or not the conclusions reached by the simulation studies still hold in actual practice.

Statement of the Problem

Previous research in statistical artifacts examined how artifacts might mathematically and contextually correlate with each other. For instance, research on the correlated relationship between range restriction and reliability has led to continuous technical refinement in artifact correction procedures that take into account of this inter-correlation among artifacts (Le & Schmidt, 2006). However, not much empirical evidence is available to demonstrate the interdependent relationship between predictor and criterion reliabilities, and neither is there an empirical examination on the impact of this interdependence on meta-analytic results. Further, there is a lack of using real-world data to show how the analysis results from the Hunter-Schmidt procedure and the RBNL procedure compare to each other under the violation of reliability independence.

Purpose of the Study

This study aims to provide an empirical examination on the problems stated right above. Specifically, this research study will explore the differences between meta-analysis results under various degrees of reliability interdependency, such as differences in meta-analytic estimates for population effect size, observed variance, credibility intervals, confidence intervals, and heterogeneity in population. Two individual artifact correction procedures, the Hunter-Schmidt procedure and the RBNL procedure, will be compared in terms of variations in the parameter estimations under reliability interdependency. This study is also interested in answering the question whether or not the methodological advantages suggested by Monte Carlo data simulation will still hold in real world case.

Specifically, the current study is to use a real-world case (a) to explore how and to what degree that reliabilities are correlated between independent and dependent variables; (b) to empirically compare the performance of the Hunter-Schmidt artifact correction procedure versus the RBNL procedure, under the condition of inter-correlated reliabilities; (c) to compare the estimated meta-analytic results from the two artifact correction procedures against these from the bare-bones procedures in order to assess the necessity of artifact correction under the condition of correlated reliabilities; and (d) to explore the differences between conclusions from data simulation studies and these from this real-world data analysis.

Definition of Key Terms

Before diving deeper into the topic of correcting correlated measurement errors and its impact on meta-analysis results, some terms should be clarified here for readers unfamiliar with the history of artifact correction meta-analysis or its methodology.

Psychometric Meta-Analysis

Since the 1980s, continuous efforts have been devoted in developing meta-analytical procedures. A meta-analysis theory was developed that adopts some basic ideas from both psychometric theory and statistical theory in order to allow corrections for statistical artifacts, such as measurement error, range restriction, and others that attenuate the true population effect size (Callender & Osburn, 1980; Raju & Burke, 1983; Schmidt, et al., 1976; Schmidt & Hunter, 1977). A collection of methods derived from this theory became known as psychometric meta-analysis (PMA). Traditional meta-analysis differs from PMA in that the former only focuses on correcting sampling errors rather than a collection of statistical artifacts as PMA does.

Validity Generalization

Validity Generalization (VG) is a more specific case of psychometric meta-analysis (PMA). VG usually sets its focus on the generalization of the validity of a selection test (e.g., personnel selection test and college entrance exam), and the effect size of interest is more often a bivariate Pearson product-moment correlation coefficient (i.e., the validity coefficient). VG differs from PMA in that the latter focuses on a variety of effect sizes and can be more broadly referred in other research fields other than just in educational and psychological studies. VG and meta-analysis are not synonymous terms either, although sometimes they are used interchangeably by Industrial and Organizational psychologists because the genesis of meta-analysis in I/O psychology is in the sphere of personnel selection validities.

Artifacts

Statistical artifacts are evidenced as the main cause of variation across individual studies in PMA (Schmidt et al., 1976). Specifically, those artifacts include but are not limited to random sampling error, unreliability caused by uncertainty in human subjects, as well as restrictions in study design, and data from samples that are less than representative of the research population or are due to inappropriate adoption of measurement tools that assess the research construct inaccurately. In addition to these, reporting or transcriptional error is considered as one artifact that is difficult to correct systematically (Hunter & Schmidt, 2004). In most cases, population estimates are considered being attenuated by various statistical artifacts that exist to different degrees of severity across individual studies.

Artifact correction procedure. The correlation-based artifact correction was built upon the theory that artifacts attenuate the true correlation coefficient by a multiplicative fraction (Schmidt et al., 1976). The major goal of PMA analysis is the estimation of the mean (M_ρ) and variance (V_ρ) of true validities (correlation coefficient) across populations by removing the effect of statistical artifacts. Briefly speaking, the correlation-based artifact correction procedures are developed on the basis of (Raju, Pappas, & Williams, 1989)

$$r = \rho abc + e, \tag{1}$$

where $a = \sqrt{\rho_{XX}}$, $b = \sqrt{\rho_{YY}}$, and $c = \dfrac{u}{\sqrt{1+(u^2-1)\rho^2 a^2 b^2}}$; r is observed correlation between predictor X and criterion Y; ρ is population correlation between X and Y and it is unbiased from measurement error and range restriction; ρ_{XX} and ρ_{YY} are unbiased population reliability of X and Y, respectively; u is the ratio of restricted population standard deviation to unrestricted population standard deviation on the predictor; e is sampling error, assuming an uncorrelated relationship between e and any of other parameters.

Reliability

The definition of reliability in classical test theory (CTT) is based on the premise that the variance in observed scores (X) is due in part to true differences in the latent trait being measured (T) and in part to error (e). This can be represented in the equation of $X = T + e$, where e is a variable that is normally distributed with a mean of zero and is uncorrelated with *T* and uncorrelated with *e* obtained from other measures (Allen & Yen, 1979). In CTT, a reliability coefficient (ρ_{XX}) is the proportion of observed score (*X*) variance accounted for by the true score (*T*) variance, as shown by the equation of ρ_{XX}= Var (*T*)/Var(*X*) (Graham, 2006). Bonett (2010) described reliability of a measure as the sum or average of *q* “parts” of this measure which could be raters (the inter-rater reliability), occasions and alternative forms (test-retest reliability), or test items (split half reliability and internal consistency), assuming these *q* parts are independent of each other.

Summary

The chapter introduced the topic of artifact correction for correlated reliabilities, outlined the two artifact correction approaches, and stated some specific research problems and questions to be investigated in the current study. Some basic concepts of artifact correction meta-analysis were also briefly described as well in order to prepare readers who are not familiar with these terms. This chapter sets a foundation of this dissertation and provides readers a basic understanding of the background of this research.

CHAPTER TWO

LITERATURE REVIEW

This chapter begins with a review of the needs in correction for measurement error in meta-analysis. This review provides a background context for the practical obstacles faced by meta-analysts. It also focuses on the criticalness of using reliability estimates to correct for measurement error and the impact of inter-correlated relationship between predictor reliability and outcome reliability on meta-analysis estimations. The following section reviews two artifact correction approaches, which leads up to the next section which is a review of two specific analysis procedures. Finally, the last section addresses the importance of using real-world empirical data for comparative analysis.

Correcting for Measurement Error

From the technical perspective, having reliable scores in research is a necessary condition to ensure decent estimates for validity. Because no measurement procedure is absolutely error free, therefore the observed relationship between two specific measures underestimates the true relationship between the two constructs (Schmidt, Le, & Illies, 2003). Actually, the maximum estimate of a correlation between two variables is confined by the product of the square root of each measure's reliability (Hunter & Schmidt, 2004, p. 96; McNemar, 1969). In other words, reliability serves as the upper limit of effect size (Crocker & Algina, 1986; Worthen, White, Fan, & Sudweeks, 1999). Because of this, it would be misleading to conclude raw correlations without taking into account of measurement reliability of each variable (Behrens, 1997).

However, we may have to be careful of whether substantive scientific judgments should be based on corrections that employ an untenably low reliability for a criterion which in turn yields a much enhanced estimate of corrected correlation (Murphy, 2003). In fact, when a study reports an abnormally low reliability, a lot researchers might have considered the option of not including this single study into a meta-analysis where correction for measurement error is applied (Murphy, 2003).

Mixed Types of Reliability

When correcting for measurement error in a meta-analysis study, the form of reliability estimates (e.g., internal consistency, inter-rater) varies from study to study, and sometimes even varies within study (e.g., correcting the predictor for internal inconsistency and the criterion for inter-rater unreliability). Just like an overcorrection on effect size may occur due to correcting substantially low reliabilities in a study, using mixed types of reliability estimates in one meta-analysis for correcting for measurement error could be problematic as well. Different types of reliability estimates technically reflect different sources of measurement error (Schmidt et al., 2003). Substantial literature in personnel selection deals with the choice of performance measure (Murphy & Cleveland, 1995; Schmidt & Hunter 1996; Schmidt, Viswesvaran, & Ones, 2000). Murphy (2003) in particular pointed out that the choice of reliability estimates for a performance measure has a large impact on the conclusions drawn from research involved with this measure. He stated that "It makes little conceptual sense to correct one set of correlation coefficients for the lack of generalizability across raters, another set of correlation coefficients for the lack of generalizability across peers, and yet another set of correlation coefficients for the lack of agreement between supervisory and objective measures of performance" (Murphy, 2003, p. 396).

Correlated Reliability

In a related vein, an interdependent relationship between reliability estimates of *X* and *Y* in a meta-analysis could potentially bias the estimation results as well, since most formulae and procedures for artifact correction treat reliabilities of variables as if they were independent of each other. Theoretically speaking, when the true relationship exists between the constructs of two variables in a given study—that is, *X* and *Y* share true score variances—the reliabilities of the two measures connect based on the CTT definition for reliability which is the amount of true score variance in observed variance. Secondly, when the measurement of *X* and *Y* are taking place at the same time in the same setting, it is very likely that situational factors, such as stress, fatigue, length of the survey battery etc., create similar impact on how participants answer questions or get assessed for both measures.

Köhler et al. (2015) put the assumption of independence for reliability under an empirical investigation, in the context of Perceived Organizational Support (POS) and its antecedents and outcomes. They collected reliability information for these variables from more than 350 studies and their findings denied the tenability of reliability independence between *X* and *Y*. For example, when the criterion reliability is estimated by other-rated inter-rater agreement and POS's reliability is based on self-rated internal consistency, the Pearson correlation between the two types of reliabilities across studies was −.45; yet when both reliability estimates are self-rated coefficient alpha, this strength of this relationship decreased to .16. However, there was not any specific explanation for this significant change. They also addressed the fact that reliability estimates not only vary from study to study, but also vary within a study (the types of reliability estimates for *X* and *Y* are not always the same across studies or within a study), and this will lead the correction for measurement error varying accordingly in meta-analysis. The degree of inter-dependence among different types of reliabilities would result in either overestimation or underestimation of population correlations depending on the direction of the relationship between the reliabilities. However, they did admit that "We do not know whether the correlation between reliabilities and our failure to account for their interdependence actually leads to a substantial bias in meta-analytically derived corrected mean correlation coefficients" (p. 376).

Two Types of Reliability Estimates

Reliability is conventionally defined in CTT as the square of the correlation between observed scores and true scores ρ_{XX}= Var (T)/Var(X) with X=T+e, which is very conceptual in that the true scores (T) are always unknown. This definition leads directly to the formula for attenuation correction that allows for estimating the correlation among true scores (Murphy & DeShon, 2000). The definition of error (e), a variable that is assumed uncorrelated with either T or with other e, is essential to this theory. Unless this assumption is met, the correction for attenuation cannot provide a valid estimate of true score correlation and will lead to a biased meta-analytical conclusion.

Internal consistency. As the most commonly used reliability estimate for educational and psychological measures (Bollen, 1989, p. 215), coefficient alpha is an index of internal consistency and reflects the degree of item agreement within a measure. Recent research from Schmidt et al. (2003) indicated that there is transient measurement error in self-reported psychological measures that often gets ignored (e.g., personality and ability measures). Transient error occurs across occasions and is caused by variations in subject's mood, feeling, mental efficiency, or mental state across occasions (Schmidt et al., 2000). Failing to capture this type of error substantially attenuates observed relationships between constructs and leads to erroneous conclusions unless it is appropriately corrected for. This implies that the common practice of using coefficient alpha for correcting unreliability in meta-analysis typically leads to under-correction and potentially produces a downward bias in estimates of population mean correlation. This type of reliability does not detect or remove the effects of transient error since it is estimated through self-reported data collected at one point in time, and it only captures the random response error which occurs across items within the same occasion, as well as specific factor error (e.g., when one has dyslexia issues with reading survey items).

Intra-rater reliability. In personnel selection, it is quite common that job performance or organizational behaviors are rated by supervisors and coworkers. Just as in psychiatric therapy, clinical counselors have to evaluate patient's condition beyond patient's self-report syndromes. However, Ones, Viswesvaran, and Schmidt (2008) argued that the most appropriate reliability coefficient for performance or behavior ratings for those studies should be inter-rater reliability estimated by the correlation between ratings produced by different raters at Time 1 and Time 2 across a reasonable time interval. However, in usual cases it is more commonly seen that only one rater is assigned to one ratee at a single time point and it is not the same rater for all ratees. Reliability estimated in these common cases is intra-rater reliability (indexed as Conbach's alpha). Note that one should never use a mixture of intra- and inter-rater reliabilities in the artifact-distribution-based correction since the sources of measure error are different (Hunter & Schmidt, p. 137). Particularly, the use of intra-rater reliabilities is claimed to lead to very severe under-corrections for measurement error, since it does not capture either transient error or rater bias due to rater-ratee interactions (Schmidt et al., 2000).

Two Types of Artifact Correction Approaches

Individual Correction

When information is available for the size and nature of the artifacts in most studies, and the few missing values can be replaced by the mean value across the studies where information is given, then each observed correlation is able to be corrected individually and the meta-analysis could be conducted on the corrected correlations. This type of artifact correction in meta-analysis refers to individual artifact correction. In other words, if the reliability of each variable is known in individual study, error of measurement can be eliminated from a meta-analysis at the level of single study.

Artifact-distribution-based Correction

However, when information for artifacts is only sporadically available, artifact distributions (i.e., mean and variance for each artifact) is suggested being used to correct the observed effect size distribution at meta-analysis level. This usage of artifact distribution to correct effect size distribution refers to artifact-distribution-based correction or distributional meta-analysis (Murphy, 2003). If correction is done this way, the accuracy of meta-analytic estimates functionally depends on artifact distributional parameters. Sometimes, the values used in artifact distributions can be taken directly from the studies that contribute correlation coefficients to the meta-analysis. In other cases, however, the information of artifacts from the studies is so sparse that the artifact distributions have to be hypothetically assumed based on researchers' knowledge in the research topic, and this is specifically called hypothetical artifact distribution (Hunter & Schmidt, 2004).

Challenges for the Two Approaches

Meta-analysis studies that use artifact-distribution-based approach tend to generate population estimates that are not quite as accurate as those from the studies where the approach of correcting coefficient individually is applied (Hunter & Schmidt, 2004, p. 157). The conundrum in reality is that the needed information for correcting artifacts in each individual study is often not available, thus distributional meta-analysis must be used if one still wishes for a correction in this situation (Hunter &Schmidt, 2004, p. 167).

However, the application of artifact-distribution-based correction can be more challenging than individual correction. In the artifact correction meta-analysis world, there are five classic artifact-distribution-based correction procedures including the interactive procedure (Pearlman, Schmidt, & Hunter, 1980), the non-interactive procedure (Schmidt, Gast-Rosenberg, & Hunter, 1980), the independent multiplicative procedure (Callender & Osburn, 1980), and the Taylor Series Approximation 1 (TSA1) and TSA2 procedures (Raju & Burke, 1983). Burke and Landis (Murphy, 2003, p. 287) discussed five methodological and conceptual challenges using these five procedures to make extrapolations about relations of constructs and the effectiveness of behavioral intervention. They pointed out that the accuracy of estimates yielded by the five well-known correction procedures substantially depend on the degree of congruence between the hypothetically assumed artifact distribution and the true distribution (Paese & Switzer, 1988; Raju et al., 1989). All five procedures commonly adopted a hypothetical artifact distribution presented in Pearlman, Schmidt and Hunter (1980), and this distribution was only meant for personnel selection test settings. When an indiscreet adoption of this distribution occurs in domains other than personnel selection, questions will arise around the accuracy and effectiveness of these artifact correction procedures. In addition, a meta-analysis study in a new research domain or topic where not much of literature is available will have no reliable source to refer to for a hypothetical artifact distribution. In conclusion, the merit of artifact-distribution-based correction would be unknown due to the impossibility of knowing the extent to which the assumed artifact distribution represents the unknown population artifact distribution.

Another apparent shortcoming of using artifact distribution is that it automatically assumes homogeneity of reliabilities and it does not differentiate the forms of reliability. Reliability can be estimated via multiple procedures to capture specific sources of measurement error and can result in different sampling distributions. Awareness of such variety is important for researchers to reach more judicious conclusions. When general conclusions on measurement error are made through combining different types of reliability, a combined estimate of reliability at the meta-analytic level will confound the sources of measurement error captured by each type. Transforming the reliability estimates into one form might be an alternative solution. However, the problem is that a transformed coefficient alpha displays an *F* distribution (Feldt, 1965; Kristof, 1963), and split-half correlations are typically adjusted for test length, which results in a sampling distribution different from that of the unadjusted split-half correlation (Lord, 1974). Therefore, reliability transformation might not be the ultimate solution. But the other way we can explore this is that upon a sufficient amount of reliability estimates, studies with the same type of reliability can be synthesized in a group, separated from those that reported a different form of reliability (Beretvas & Pastor, 2003; Sawilowsky, 2000).

Thirdly, when artifact-distribution-based approach is in use, it generally requires a certain degree of independence among artifacts across studies (i.e., the pair-wise independence assumption). This assumption of artifact independency across studies means the artifacts are independent of each other and independent of the size of the true population correlation (Hunter & Schmidt, 2004; Raju et al., 1998), and this applies to most distribution-based correction procedures, such as the five classic ones mentioned earlier. One of the pair-wise independence assumptions is that measurement error of X is uncorrelated with measurement error of Y across studies. However, empirical evidences have shown that this assumption is likely to be challenged in practice (Köhler et al., 2015; Mendoza & Reinhart, 1991).

Individual correction on the other hand avoids all the three issues discussed above which artifact-distribution-based correction has to face. Different sources of measurement error can be corrected individually for each study, and then meta-analysis is conducted based on those corrected sample correlations. Therefore, the only homogeneity we should focus on is the homogeneity of sample correlations, and this approach does not necessarily require independence among artifacts across studies since artifacts have already been corrected at individual level.

However, the limitation of individual correction cannot be avoided when reliability estimates correlate within a study. This independence assumption for artifacts within a study means that measurement error of X should be uncorrelated with measurement error of Y in a given study. While this assumption has been theoretically and empirically contested by an opposite finding from Köhler et al. (2015), the impact of this relationship on artifact correction has not been addressed sufficiently (Raju et al., 1998; Zimmerman & Williams, 1977). Luckily, the artifact correction meta-analysis method developed by Raju et al. (Raju et al., 1991; Raju & Brand, 2003) has been specifically addressed to be applicable for any definition of reliability

(Raju & Brand, 2003). In this regard, the current research will expect the RBNL procedure to provide more accurate meta-analytic estimates than those from other artifact-correction procedures.

Two Individual Correction Procedures

The RBNL Procedure

The RBNL procedure essentially adopts a sample-based individual correction approach, therefore it naturally suffers when artifact data is partially available or hypothetical artifact values are used to replace missing values (Murphy, 2003, p. 273). However, the major advantage of the RBNL procedure is that it takes account of sampling errors of the artifacts themselves (Raju et al., 1991; Raju & Burke, 2003). In addition, it does not explicitly require one of the main assumptions underlying most of the correlation-based procedures which is that the population correlation *(p),* predictor reliability, criterion reliability, and range restriction are uncorrelated across populations and within a population (Raju & Burke, 1983; Raju et al., 1989; Raju et al., 1998). Thirdly, the RBNL procedure is stated as a suitable artifact correction method regardless of the type of reliability estimates because of its ability of accounting for the sampling errors of artifacts. An overview of the RBNL procedure is summarized below.

The theoretical foundation of the RBNL procedure starts with

$$\hat{\rho}_i = \rho_i + e_i, \tag{2}$$

where ρ_i is the unrestricted and unattenuated population correlation for sample (study) *i*; $\hat{\rho}_i$ is an estimate of the unattenuated and unrestricted population correlation ρ_i for sample (study) *i*; e_i is the sampling error associated with $\hat{\rho}_i$. It is assumed that the expectation of e_i is zero and there is an independent relationship between ρ_i and e_i across populations. In fact these two assumptions may be considered tenable in most empirical studies (Raju et al., 1989), which lead to the equation of

$$M_\rho = M_{\hat{\rho}} \tag{3}$$

and

$$V_\rho = V_{\hat{\rho}} - V_e . \tag{4}$$

In order to obtain M_ρ, the first step of the RBNL procedure is to calculate $\widehat{\rho}_i$ for each individual study *i* using Equation 5 which is derived based on classical test theory (Lord & Novick, 1968).

$$\widehat{\rho}_i = \frac{k_i r_i}{\sqrt{r_{X_i X_i} r_{Y_i Y_i} - r_i^2 + k_i^2 r_i^2}}, \tag{5}$$

where r_i is the correlation between the predictor *X* and criterion *Y* in a sample (study) *i* from the attenuated and restricted population ρ_i, and it is a simpler expression of $r_{X_i Y_i}$; $r_{X_i X_i}$ is the sample-based, observed reliability for predictor *X*; $r_{Y_i Y_i}$ is the sample-based, observed reliability for criterion *Y*; u_i is the ratio of SD_i over SD_a, the ratio of restricted standard deviation (SD_i) to unrestricted standard deviation (SD_a) on *X*; k_i is equal to $1/u_i$. After each $\widehat{\rho}_i$ is calculated from observed values according Equation 5; $M_{\hat{\rho}}$ can be obtained through the sample size weighted average of $\widehat{\rho}_i$ and ready to be used as an estimate of M_ρ based on Equation 3.

In order to estimate V_{ρ} from Equation 4, one should first calculate $V_{\hat{\rho}}$ the variance of $\widehat{\rho}_i$ from the set of sample studies. V_e, the sample-size weighted average of $\hat{V}_{e_i}$, can be calculated following Equation 6.

$$\hat{V}_{e_i} = \frac{k_i^2 r_{X_iX_i} r_{Y_iY_i} (r_{Y_iY_i} - r_i^2)(r_{X_iX_i} - r_i^2)}{(N_i - 1)\hat{W}_i^3}, \tag{6}$$

where $\hat{W}_i = r_{X_iX_i} r_{Y_iY_i} - r_i^2 + k_i^2 r_i^2$. The Equation 6 forms the basis for calculating $\hat{V}_{e_i}$. However, this calculation may vary depending on how much information for artifacts is reported in a sample study, and the following section provides detailed calculation procedures for seven different scenarios in VG research.

(1) When u_i is unavailable: one possible substitute for each missing u_i is the weighted average of all available range restriction values in a given set of studies. A suggested alternative is to use the hypothetical distribution of range restriction recommended by Schmidt, Hunter, and their colleagues (Pearlman et al., 1980) or by Alexander, Carson, Alliger and Cronshaw (1989) particularly for VG studies. Both have been popularly used in VG research, but the second option is preferred when none of the validity studies in a given meta-analysis provide information for range restriction. With an estimate for u_i, one can readily use Equation 5 and Equation 6 to estimate $\widehat{\rho}_i$ and its sampling variance, and then proceed with the rest of the meta-analysis as previously outlined.

(2) When $r_{X_iX_i}$ is unavailable: using weighted mean of all available predictor reliability or the recommend hypothetical distribution for predictor reliability (i.e., an average

reliability of .80) by Pearlman et al. (1980), one can estimate $\widehat{\rho}_i$ and its sampling variance using Equation 5 and Equation 7.

$$\hat{V}_{e_i} = \frac{k_i^2 r_{X_iX_i}^2 r_{Y_iY_i} (r_{Y Y_i} - r_i^2)(1 - r_i^2)}{(N_i - 1)\hat{W}_i^3} \tag{7}$$

(3) When $r_{Y_iY_i}$ is unavailable: using weighted mean of all available criterion reliability or the recommended hypothetical distribution for criterion reliability (i.e., an average reliability of .60 in Pearlman et al., 1980) if it is appropriate, one can estimate $\widehat{\rho}_i$ and its sampling variance using Equation 5 and Equation 8.

$$\hat{V}_{e_i} = \frac{k_i^2 r_{X_iX_i} r_{Y_iY_i}^2 (1 - r_i^2)(r_{X_iX_i} - r_i^2)}{(N_i - 1)\hat{W}_i^3} \tag{8}$$

(4) When u_i and $r_{X_iX_i}$ are unavailable: combining situation of 1 and 2, one can estimate $\widehat{\rho}_i$ and its sampling variance using Equation 5 and Equation 7.

(5) When u_i and $r_{Y_iY_i}$ are unavailable: combining situation of 1 and 3, one can estimate $\widehat{\rho}_i$ and its sampling variance using Equation 5 and Equation 8.

(6) When $r_{X_iX_i}$ and $r_{Y_iY_i}$ are unavailable: combining situation of 2 and 3, one can estimate $\widehat{\rho}_i$ and its sampling variance using Equation 5 and Equation 9.

$$\hat{V}_{e_i} = \frac{k_i^2 r_{X_iX_i}^2 r_{Y_iY_i}^2 (1 - r_i^2)^2}{(N_i - 1)\hat{W}_i^3} \tag{9}$$

(7) When u_i $r_{X_iX_i}$ and $r_{Y_iY_i}$ are unavailable: combining situation of 6 and 1, one can estimate $\widehat{\rho}_i$ and its sampling variance using Equation 5 and Equation 9.

The RBNL procedure, which is an individual correction method, should be able to generate more accurate population parameter estimates than artifact-distribution-based correction procedures. Investigations through Monte Carlo simulations demonstrated that TSA1 was generally superior to the other four classic artifact-distribution-based procedures mentioned earlier when estimating the mean (M_ρ) and variance (V_ρ) of true validity distribution (Mendoza & Reinhardt, 1991; Raju & Burke, 1983). Although only few VG models specifically address the correlated relationship between artifacts, the RBNL procedure provides more accurate estimates when compared to TSA1 under different degrees of inter-correlation among artifacts (Raju et al., 1998). A series of Monte Carlo studies by Raju et al. (1998) proved that both the TSA1 and the RBNL procedures provide comparable estimates for the mean true validity (M_ρ), but the RBNL procedure offers more accurate estimation for the true validity variance (V_ρ).

Besides offering more accurate population estimates, the RBNL procedure provides artifact correction at an individual study level, which in turn presents researchers an opportunity to conduct statistical significance tests on corrected correlations between individual studies. Another convenience is that instead of offering a separate sampling variance formula for each definition of reliability, the RBNL procedure proposes a single formula for estimating sampling variance of corrected correlations that can be used for any form of reliability.

The Hunter-Schmidt Procedure

Hunter-Schmidt's artifact correction procedures (Hunter & Schmidt, 1990; 2004) have been the most commonly used meta-analysis procedures in applied psychology (Aguinis et al., 2011; Murphy, 2003). Frank Schmidt and John Hunter are pioneers of psychometric meta-analysis (PMA) which addresses the effect of statistical artifacts on meta-analysis estimates (Callender & Osburn, 1980; Raju & Burke, 1983; Schmidt & Hunter, 1977). The main focus of Hunter-Schmidt's procedures is on artifact correction particularly when range restriction is involved (Law, Schmidt, & Hunter 1994; Hunter et al., 2006). Yet, the reality is that as much as the information availability has been improved in social science studies, getting information for range restriction is still very challenging (Aguinis et al., 2011; Aytug et al., 2012;), and the judgment call for direct or indirect range restriction is not always very intuitive either. The practical usage of Hunter-Schmidt procedures was examined by Aguinis et al. (2011) through a collection of published meta-analysis studies in the VG field from the past 30 years. Their work revealed that a vast majority of these published studies did choose Hunter-Schmidt procedures but did not actually follow the complete analysis procedure to identify and incorporate range restriction correction. In addition, the use of Hunter-Schmidt meta-analysis procedures could potentially be limited in practice due to the requirement of the fundamental assumption that the true validity and artifacts have to be pair-wise independent from each other across populations.

If readers are interested in the specifics, Chapter Three of Hunter & Schmidt's meta-analysis book (2004) detailed the theoretical foundation and the procedure of correcting measurement errors at the level of single study for meta-analysis (Hunter & Schmidt, 2004, p. 95). The following section thus presents a brief summary of the Hunter-Schmidt sample-based individual correction procedure.

First of all, a different format was used in Hunter and Schmidt's book to denote Equation 1 as

$$r_i = \rho_i A_i + e_i, \tag{10}$$

where A_i represents the artifact attenuation factor for each study i and A_i actually is a multiplicative factor of all the correctable artifacts for a study. Therefore when $A_i = \sqrt{r_{X_iX_i}}\sqrt{r_{Y_iY_i}}$ is introduced to Equation 10, meaning only measurement errors in X and Y are taken into consideration for artifact correction, the observed correlation can be expressed as

$$r_i = \rho_i \sqrt{r_{X_iX_i}}\sqrt{r_{Y_iY_i}} + e_{r_i}, \tag{11}$$

where all the definitions are the same as described in the section of the RBNL procedure above, except e_{r_i} is denoted to represent the sampling error associated with r_i .

The following describes an individual artifact correction approach adopted by the Hunter-Schmidt procedure for estimating M_ρ and V_ρ . First, the corrected correlation is derived from

$$\hat{\rho}_i = \frac{r_i}{A_i} \tag{12}$$

for each study. Second, the sampling error variance in the uncorrected correlation r_i needs to be computed first via

$$\upsilon_{e_{ri}} = \frac{(1-\bar{r}_o^{\,2})^2}{N_i - 1}, \tag{13}$$

where $\bar{r}_o$ is the sample size weighted average of the uncorrected correlation r_i across studies. Until now, five numbers are recorded for each study including the uncorrected correlation r_i, the corrected correlation $\hat{\rho}_i$, the sample size N_i, the compound attenuation factor A_i, and the sampling error variance in the uncorrected correlation $\upsilon_{e_{ri}}$.

Sequentially, the corrected population average M_ρ can be estimated through

$$M_{\hat{\rho}} = \frac{\sum w_i \hat{\rho}_i}{\sum w_i}, \tag{14}$$

where $w_i = N_i A_i^2$. In order to estimate V_ρ through

$$V_\rho = V_{\hat{\rho}_i} - Ave(\upsilon_{e_{\hat{\rho}i}}) \tag{15}$$

the variance of corrected correlation has to be calculated first via

$$V_{\hat{\rho}_i} = \frac{\sum w_i (\hat{\rho}_i - M_{\hat{\rho}})^2}{\sum w_i} \tag{16}$$

following with the sampling error variance in the corrected correlation computed from

$$\upsilon_{e_{\hat{\rho}i}} = \frac{\upsilon_{e_{ri}}}{A_i^2} \tag{17}$$

and

$$Ave(\upsilon_{e_{\hat{\rho}i}}) = \frac{\sum w_i \upsilon_{e_{\hat{\rho}i}}}{\sum w_i}. \tag{18}$$

Hunter and Schmidt stated that error of measurement should always be corrected especially when conducting meta-analysis studies, regardless of the fact that reliability estimates used for correction are not always accurate and there is possible a lack of information for the effect of range restriction on reliabilities (Hunter & Schmidt, 2004, p. 103). But they did not provide any empirical evidence to address whether this conclusion would still hold when a severe interdependence between reliabilities is present. Fortunately, the current study sets its effort in investigating the effectiveness of the Hunter-Schmidt correction for correlated measurement errors.

The Bare-Bones Procedures

When sampling error is the only artifact to be considered, the bare-bones meta-analysis model can be used. Because of its simplicity, the bare-bones model has actually been widely used in a variety of different disciplines (e.g., psychology, ecology, engineering, medicine, pharmaceutics, epidemiology, business, and consumer research) where measurement reliabilities and range restrictions are not commonly recognized.

Methods for conducting this kind of random-effects meta-analysis model has been evolving just like many other meta-analytical techniques. Two bare-bones procedures have been widely used, the Hedges-Vevea procedure (Hedges & Vevea, 1998) and the Hunter-Schmidt bare-bones procedure (Hunter & Schmidt, 2004; Schmidt & Hunter, 1977). The latter one is often found widely applied in the Industrial and Organizational field as opposed to the Hedges-Vevea procedure which is much more popular in other research fields. The major differences between these two procedures reside in whether or not the Fisher's z-transformation for observed correlation effect size r is used, as well as the weighting scheme in estimating population correlation effect size ρ based on observed value of r. The Hedges-Vevea procedure first transforms observed r to Fisher's z and then conducts the proceeding analyses based on z values, whereas the Hunter-Schmidt procedure uses observed r value throughout the computation. The Hedges-Vevea procedure uses an optimal weighting scheme where both the estimated population variance and the study sample size are considered, while the Hunter-Schmidt method weights observed r using its associated sample sizes. Few Monte Carlo studies showed that both of the bare-bones procedures generated quite comparable estimation results, and some studies even detailed comparison between these two procedures (Field, 2001; Hall & Brannick, 2002).

Estimates from the bare-bones model will be used in the current study as a baseline benchmark in contrast to the estimates from other artifact correction models. In this way, any significant differences attributable to artifact correction procedures can be easily red-flagged (Murphy, 2003, p. 359; p. 405). Murphy (2003) advocated for research to return to the bare-bones methods instead of overusing and overdeveloping artifact correction procedures. He stated that all the complexity in artifact correction models did not actually produce much difference among them in terms of estimation results, whereas results from the bare-bones meta-analysis turned out to have minor differences with those that involved artifact correction.

Admittedly, prevalent opinions on the weakness of the bare-bones model being used in social science is that it takes only the sampling variance into account but no other artifacts, which would inevitably introduce a bias toward overestimation of the true correlation variance (Callender, Osburn, Greener, & Ashworth, 1982). Studies based on data simulation indicated a quite pronounced tendency for the bare-bones model to overestimate the true variance and to underestimate the lower credibility value (Callender et al., 1982). This tendency would lead to an under-generalization of validity. However, when reliability interdependence is present in a real-world case, it would be interesting to see if the population estimations between the bare-bones estimates and these from artifact correction procedures are significantly different.

Using Real-World Data

The current study proposes an empirical investigation instead of a theoretical testing. Myriad Monte Carlo studies have been done examining the precision and bias of various meta-analytic procedures, but it is questionable that these conclusions reached by the simulation data still hold true in actual practice. Monte Carlo data simulation methods prevail in testing how meta-analysis results vary under different degrees of violation of the underlined assumptions, and addressing the strength and weakness of each method and procedure (Callender & Osburn, 1981; Hall & Brannick, 2002; Law et al., 1994; Oswald & Johnson, 1998; Raju et al., 1998; Vacha-Haase, 1998; Schmidt, Law, Hunter, Rothstein, Pearlman, & McDaniel, 1993). In this sense, Monte Carlo studies can be extremely useful for studies focusing on methodology comparison in that they often provide important insights into the similarity and difference among various analytic approaches.

However, the central weakness of most Monte Carlo studies resides in the gap between using the simulated data for research methodology development and the actual research in practice where much more complexity exists in terms of assessing research assumptions. Therefore, an empirical exploration in the effect of reliability interdependency using real-world data is needed, not only to bridge the gap between data simulation and actual application in the field practice, but also to provide some basic understanding on how correlated reliability estimates affect parameter estimates in meta-analysis.

Summary

This chapter reviewed some challenges and issues which artifact correction meta-analysis has been facing, particularly for correcting measurement error and using artifact-distribution-based meta-analysis. Correcting measurement error using individual artifact correction approach is hence promoted in this study. Two individual correction procedures were reviewed in this chapter. Built on those concepts, the next chapter will present the study design and analysis method in order to evaluate the empirical performance of the two artifact correction procedures under reliability interdependence.

CHAPTER THREE

METHODS

The purpose of this chapter is to present the methods that were used to evaluate the research goals set in Chapter One. The first section introduces the procedure of collecting and coding selected studies from which information was extracted. The second section presents the basic study characteristics, the magnitude of correlations between reliability estimates, and different statistical indices across four sets of data analyzed in the current study. The last section describes the methods of meta-analysis (i.e., the Hedges-Vevea procedure, the Hunter-Schmidt bare-bones procedure, and the RBNL artifact correction procedure) that were compared using the data collected for this study. Specifically, the main focus of the last section is on how the estimation results from the different procedures compared to each other, when taking into account the types of reliabilities and the magnitude of the correlation between reliabilities.

Selection and Inclusion of Studies

Data of the current study were extracted from a subset of studies from Köhler, Cortina, Kurtessis and Gölz's research (2015). Köhler et al. searched and identified a large collection of studies involved Perceived Organizational Support (POS) and variables such as Organizational Affective Commitment, Job Satisfaction, Organizational Citizenship Behavior (OCB), and Job Performance (JP). POS is defined as the employees' perception about the degree to which the organization appreciates their contribution and would treat them favorably or unfavorably in diverse situations (Eisenberger, Fasolo, LaMastro. 1990). JP is defined as a fulfillment of tasks that are required by the formal job description. OCB is considered as behaviors that arenot only beneficial to the organization but also go beyond formal job requirements, such as helping co-workers at work, working extra hours, making suggestions for improvement, and other related behaviors (Borman & Motowidlo, 1997). OCB also includes actions favorable to the organization as a whole (OCB to Organization) and those favorite to other individuals such as supervisors and coworkers (OCB to Individuals) that go beyond assigned responsibility. The reliability information for variables of POS and its antecedents and outcomes was gathered by the researchers to calculate the extent to which these reliability estimates correlate.

Köhler et al.'s literature search was conducted with rigor. Both published and unpublished citations were quested between the timeframe of 1986 to 2011 in ABI/INFORM, APA PsycNET, PsycINFO, ProQuest Research Library, Digital Dissertations, Google Scholar, and the Defense Technical Information Center using key words *perceived organizational support, organizational support, perceived support*, and *POS*. Previous meta-analyses studies on POS were also used as the stems to track down any references that were not identified in the first round of search. There were 345 citations included in their study containing critical information that could be used for the current research purpose (e.g., sample size, reliability coefficients, correlation coefficients, and other study descriptions). A particular advantage of adopting the studies they gathered is that the unidimensionality of the construct of POS was already ensured by Köhler et al.'s searching procedure. This is because Köhler et al. set up a key inclusion criterion that a study can only be included if the common measure of POS developed by Eisenberger, Huntington, Hutchison, and Sowa (1986) was used in that study.

To meet the need of the current study, only the articles reporting the Pearson correlation coefficient between POS and one of the two criterion variables JP or OCB were included. This inclusion criterion was put into place due to the consideration that JP and OCB represent two independent yet common aspects of Outcome Performance which has been studied as one of the key consequences of POS (Rhoades & Eisenberger, 2002). In addition, both internal consistency and intra-rater reliability estimates were collected respectively for JP and OCB with a decent amount of studies as a result of Köhler et al.'s research. Information of the two types of reliability estimates facilitated explorations on how the interaction between the type of reliability and the magnitude of correlation between reliability estimates impact artifact correction results, particularly when using different meta-analysis procedures. However, because OCB was measured and reported with high diversity and complexity across these sample studies, the current research therefore further recoded and analyzed the data separately from the way that OCB scores were reported. For instance, three separate datasets were created for OCB-Overall, OCB to Organization (OCB-O), and OCB to Individuals (OCB-I), respectively. This data organization aimed to maximize the integrity of the measurement construct for each variable involved in the current study, and to ensure that studies included in each data set were theoretically and rationally similar to each other. This separation for outcome variables of OCB-Overall, OCB-I, and OCB-O might reduce the number of studies included in each dataset but increase the interpretability of the analysis results.

Out of the total 345 articles Köhler et al. cited in their reference list, there were 340 articles successfully retrieved and reviewed. The current research only included studies that contained full or partial information that can be used for artifact correction meta-analysis for the correlation of POS with JP, the correlation of POS with OCB-Overall, the correlation of POS with OCB-I and, last but not the least, the correlation of POS with OCB-O. This restriction excluded 207 articles and left 132 articles to be included in the present study.

Study Design

Table 1 presents the design of the study analysis and how the data extracted from the collection of studies are arranged for the current analysis needs. This research essentially involved a three-factor design with the method of analysis as one of the factors. There are four sets of data examining the correlations of POS and its four criterion measures, as well as two types of reliability estimates (self-rated vs. other-rated) for each criterion variable, together generating eight combinations. Each combination was analyzed by four meta-analysis procedures including the Hedges-Vevea procedure, the Hunter-Schmidt bare-bones procedure, the Hunter-Schmidt individual artifact correction procedure, and the RBNL individual artifact correction procedure.

Table 1. Data organization and study design for the current research

	Set 1		Set 2		Set 3		Set 4	
Predictor variable *X*	POS		POS		POS		POS	
Reliability of *X*	(internal consistency)		(internal consistency)		(internal consistency)		(internal consistency)	
Criterion variable *Y*	JP		OCB-Overall		OCB-I		OCB-O	
Reliability of *Y*	self-rated (internal consistency)	other-rated (intra-rater)	self-rated (internal consistency)	other-rated (intra-rater)	self-rated (internal consistency)	other-rated (intra-rater)	self-rated (internal consistency)	other-rated (intra-rater)
Number of studies	32	59	15	15	32	62	23	28

Coding and Data Compilation

Each of the 132 articles was reviewed for relevance and coded for information. No subjective coding was essentially necessary since the needed information can be easily found in the method section in most studies. At first, all the articles were reviewed and coded by the current researcher who is familiar with both the topic of Industrial and Organizational Psychology and Quantitative Research Methodology. Later, the coding results were sent to another coder who was a graduate student in Statistics for a random cross-check. Whenever coding discrepancies existed, they were resolved by a series of discussions between the two coders until a consensus was reached. The identified key information can be organized by the three main categories listed below.

Study Characteristics

This category included descriptive information about the sample studies themselves such as year of publication, whether research is funded or not, participants' occupation (e.g., clerk, mechanical technician, salesman), the type of organization (e.g., research institution, university, firms, corporates), the geographical region where the research was conducted (e.g., the name of the countries), and publication status (e.g., unpublished dissertation or thesis, journal publication, evaluation reports). In field research, information from this category potentially can be used for detecting the effect of moderators (if there is any) on the relationships between POS and its criterion variables. However, information from this category does not involve the main focuses of the current research.

Effect Sizes and Reliability

Information in this category served for detecting the correlated relationship between reliability estimates of predictor and outcome variables, as well as for artifact correction for measurement error. This category contained information of the variable names (e.g., perceived organizational support, job performance, organizational citizenship behavior), the observed bivariate correlation coefficient as effect size, the descriptive statistics for the variables of interest (e.g., sample size, mean, standard deviation or variances), the type of reliability (e.g., internal consistency, test-retest, intra-rater, or inter-rater), the value of reliability estimates, the number of items used in the instrument, and the rating scales such as the 1 to 5 or 1 to 7 Likert scale. Reliability coefficients were coded as internal consistency when respondents completed a measure about themselves at the same time point. If the target of the measures was different from the person who completed the questionnaire (e.g., peer reports, supervisor reports), reliability values for those measures were noted as intra-rater reliability and indexed with Cronbach's alpha from raters.

If a study reported bivariate correlation and reliability information on more than two criterion variables, it contributed multiple independent effect sizes. This is because these outcome variables represented different constructs. For example, if a study measured and reported correlation coefficient of POS with JP, as well as correlation of POS with OCB-Overall, even from the same group of participants, this one study still offered two independent effect sizes for the two sets of data that will be meta-analyzed separately (Table 1). Multiple effect sizes from one article can also be treated as if they were from different studies where each effect size was calculated from an independent sample and their sampling errors were not correlated. For example, Pazy and Ganzach's study (2009) contributed three independent effect sizes for the correlation of POS and JP, because these effect sizes were calculated from three different samples—salesman, students, and customer service representatives.

However, a composite correlation effect size or a composite reliability coefficient should be calculated, especially when a study used a measure that contains multiple subscales and reported information for each subscale but not for the entire measure. This is because multiple effect sizes or data points from these subscales are non-independent since they were computed from data collected from the same sample of participants, and also these subscales of one measure represent similar aspects of the overarching construct. For instance, many studies measured the same group of participants using the OCB scale developed by Organ (1988) and Podsakoff and MacKenzie (1989). These two instruments included subscales of altruism, conscientiousness, sportsmanship, and civic virtue, where the first three subscales together represent the OCB to Individuals aspect (OCB-I) and the civic virtue subscale is intended to measure the OCB to Organization aspect (OCB-O). These studies either reported scores for both overall scale level and subscale level or scores for just one of the two levels. Similar reporting patterns were observed for another set of studies that measured OCB using the instrument developed by Moorman & Blakely (1995). This instrument included subscales of helping, voice, action, and loyalty, where the helping and voice subscales represent the OCB-I aspect and action and loyalty represent the aspect of OCB-O. At other times, Williams and Anderson's (1991) instrument was used to measure OCB that benefits both the larger organization (OCB-O) as well as a specific individual (OCB-I). OCB was also measured in some other studies as part of extra-role performance (OCB-O) and contextual performance which is considered as equivalence to

OCB-I (Van Dyne, Graham, & Dienesch, 1994; Williams & Anderson, 1991). Fortunately, the definition of JP was much more consistent across the vast majority of studies as in-role performance or task performance meaning a completion of tasks as part of job duty. These studies often used a single set of items with no subscales to measure JP as one construct. In terms of POS, its measurement construct was the most unidimensional one in that all the sample studies used items from the original 36-item instrument developed by Eisenberger et al. (1986).

For those studies that collected data from same group of participants using a measure that contained multiple subscales and reported scores for each of the subscales instead of the whole measure, a single composite correlation of effect size and scale level reliability was calculated using the equations provided by Hunter & Schmidt (2004, p. 433). This step ensured effect sizes were independent of each other by including only one effect size calculated from a participant group for one set of study listed in Table 1. This is particularly important since a fundamental assumption for meta-analysis is independence between effect sizes. A violation of this assumption will underestimate the sampling error variance in the observed variance of effect sizes. A brief summary of the calculations for the composite effect size and reliability are provided below. The notations used in Equation 19 to Equation 23 are different from those used in Equation 1 to Equation 18.

In order to acquire composite effect size, Hunter & Schmidt (2004) provided an equation to describe the relationship between the average correlation effect size and the composite correlation as below. For a study, if information for predictor X is reported at the whole scale level but scores for criterion variable Y are reported only at subscale level, the composite correlation can be calculated following Equation 19.

$$r_{XY} = \frac{\bar{r}_{Xy}}{\sqrt{\frac{1 + (n-1)\bar{r}_{yy}}{n}}}, \tag{19}$$

where r_{XY} is the composite correlation between score *X* and *Y* at the whole scale level; $\bar{r}_{Xy}$ is the average correlation between the individual subscale scores y_i and the whole scale level score *X*; $\bar{r}_{yy}$ is the average correlation between the individual subscale scores y_i; *n* is the total number of subscales. If the needed information to calculate $\bar{r}_{yy}$ is not available, $\bar{r}_{Xy}$ will be used as the composite correlation effect size r_{XY}.

For a measure with multiple subscales, the Spearman-Brown formula of Equation 20 was used to calculate the composite reliability for the scale as a whole.

$$r_{YY} = \frac{n\bar{r}_{yy}}{1 + (n-1)\bar{r}_{yy}}, \tag{20}$$

where r_{YY} is the composite reliability at the scale level; $\bar{r}_{yy}$ is the average correlation between the individual subscale scores y_i; *n* is the number of subscales.

All the studies from the 132 articles successfully reported information for effect sizes and reliability estimates of POS. However, in total there were 11 missing values for the reliability of criterion variables. Two cases were missing JP internal consistency estimates, another eight cases were missing intra-rater reliability estimates for JP, and the last missing case was for the internal consistency estimate of OCB-Overall. Considering there was not a large amount of missing cases in the current database, when a study failed to provide reliability information for the relevant criterion variables, the missing value was substituted with the average reliability of the variable across all samples in that set of data.

Final Database

A total of 242 independent effect sizes were obtained from studies reported in the 132 articles. Twelve articles reported results from more than one participant sample, where nine of them reported results based on two participant samples and three articles reported results for three participant samples. Six articles reported studies including information for both supervisor ratings and self-ratings for the same criterion variable. Forty-six of the articles were unpublished reports as doctoral dissertations or master thesis papers, and the remaining eighty-six were published journal articles. Of these journal articles, there were fifteen articles authored by one person, the vast majority was published by two to three authors, and the remaining sixteen articles had four or more authors for each article. About 64% of the studies were conducted in the United States and the rest were conducted in other countries. Most of the studies were conducted with a cross-sectional design and 9% of the studies were considered as longitudinal studies. A full list of the articles that were coded in the current study are listed in Appendix A.

Data Analysis

Reliability Analysis

The distributions of the predictor and criterion reliabilities as well as a scatter plot were examined before Pearson correlation analyses on reliability estimates were conducted. For each set of data presented in Table 1, a correlation between reliability estimates of POS and the criterion variable was calculated for the combination of predictor (self-rated internal consistency) with criterion (self-rated internal consistency), and the other combination of predictor (self-rated internal consistency) with criterion (intra-rater). Adopted from Kholer et al.'s study, the sample-size weighted correlation for reliability estimates can be obtained using the following equations.

$$r_{weighted} = \frac{\sum N_i (r_{X_iX_i} - \bar{r}_{XX})(r_{Y_iY_i} - \bar{r}_{YY})}{(\sum N_i - 1) sd(r_{XX})_{weighted}\ sd(r_{YY})_{weighted}} \tag{21}$$

$$sd(r_{XX})_{weighted} = \sqrt{\frac{\sum N_i (r_{X_iX_i} - \bar{r}_{XX})^2}{\sum N_i}} \tag{22}$$

$$sd(r_{YY})_{weighted} = \sqrt{\frac{\sum N_i (r_{Y_iY_i} - \bar{r}_{YY})^2}{\sum N_i}}, \tag{23}$$

where $r_{X_iX_i}$ represents the predictor reliability coefficient from study *i*; $r_{Y_iY_i}$ represents the criterion reliability coefficient from study *i*; N_i is the sample size of study *i* from which a reliability estimate was obtained; $\bar{r}_{XX}$ is the simple average of the predictor reliability values across all studies; $\bar{r}_{YY}$ is the simple average of the criterion reliability values across all studies.

Artifact Correction

In the context of meta-analysis when the population correlations of the included studies are free of other artifacts and sampling error is the only one to be considered, then the bare-bones analysis procedures should be implemented. When following the Hunter-Schmidt bare-bones procedure (Hunter & Schmidt, 2004, p. 81), the observed effect sizes from individual studies are simply weighted by associated sample sizes in order to estimate the population effect size and the corresponding sampling variance across studies. On the other hand, when following the Hedges-Vevea's procedure (Hedges & Vevea, 1998), the observed effect sizes were first converted to Fisher's *z* scores, then the analyses were conducted using these transformed *z* scores instead of the observed correlation scores.

When the statistical artifacts of measurement error are considered and correction for unreliability for both predictor and criterion variables is wanted, both the Hunter-Schmidt correction procedure and the RBNL procedure require first to compute a compound factor A for each participating study. This factor A is a square root of the multiplication between the scale reliabilities of predictor and criterion variables from each study, and is used for correcting only for effect size attenuation. In order to estimate the true population effect size, the observed effect sizes were first corrected for attenuation by this compound artifact correction factor A for each study, and then the corrected effect sizes were weighted for differences in sample size of each study. The main difference between the Hunter-Schmidt artifact correction and the RBNL correction procedure resides in the way how the sampling variances were calculated described in Chapter Two. For each study, the Hunter-Schmidt procedure estimates the sampling variance for the corrected effect size through weighting the sampling variance for the uncorrected effect size by a multiplicative term of sample size and attenuation factor A. This method does not address the sampling errors of the artifact themselves as does the RBNL procedure. On the other hand, as indicated in Equation 6 in Chapter Two, the RBNL procedure holds account for sampling errors of the artifacts themselves and takes into consideration of the interactive relationships between reliabilities, range restriction, and effect size when estimating sampling variances for each study.

Correction for biases due to range restriction on predictor X was not considered in the current study (i.e., assuming no restriction of range exists), therefore the range restriction indicator u was set to unity for all sample studies and hence there was no sampling error in range restriction values. Range restriction occurs when estimated variances are reduced due to pre-selection or censoring of sample participants, indicating a reduced representativeness of the study sample in population. Meta-analysis results that are estimated from range-restricted samples could be biased ones. However, no explicit information for range restriction was identified among the sample studies collected at the current research, and a thorough review on past meta-analysis studies for POS yielded no fruitful results for a theoretical estimate for range restriction on POS. The correction for range restriction was hence not applicable here.

Meta-Analysis

Random-effects meta-analysis models (Hedges &Vevea, 1998; Hall & Brannick, 2002) were used for both the two bare-bones analysis procedures and the two artifact correction procedures. Random-effects models assume that the distribution of effects sizes in a population of studies has a variance due to factors (such as study context) other than sampling error. That is any effect size from an individual study is expected to have its own contextual population effect sizes ρ_i which is an estimate in theory if an infinite sample could be gathered in a study context. If one conducts a lot of studies for each of the various contexts and pools all the effect sizes together, there will be a super-population of ρ. This super-population of ρ will have a distribution depicted by M_ρ, sampling variance $V_{\rho,}$, and heterogeneity or between study variance index tau square (τ^2). τ^2 is often reported in the form of an estimate of the between-context variance among contextual population effect sizes, representing the total amount of heterogeneity among the true effects. Translated, this means that the between-context variation in true effect sizes must be due to study contexts because it cannot be attributed to statistical artifacts. The results from random-effects analyses are therefore considered more generalizable beyond the included set of studies, and their inferential results can be used for prediction for what would likely happen if a new study were conducted in any context. The primary goal of the meta-analytic random-effects model is to estimate the amount of heterogeneity (τ^2) in the super-population. The random-effects model is generally preferred by meta-analysis practitioners as opposed to the fix-effects model in which τ^2 is always assumed to be zero. Technically speaking, for the Hedges-Vevea procedure the Random Effects Variance Component (REVC) τ^2 is in the metric of z using DerSimonian-Laird estimator, while τ^2 is in the metric of *r* and specified by the Hunter-Schmidt estimator for parameter estimation for the Hunter-Schmidt bare-bones procedure, the Hunter-Schmidt individual artifact correction procedure, and the RBNL procedure (Viechtbauer, 2005). It is important to know that estimation of τ^2 can be imprecise especially when the number of studies is small. The bias in the Hunter-Schmidt estimator is more likely to be substantial than the bias induced from the DerSimonian-Laird estimator (Viechtbauer, 2005).

Population parameters. Once the population parameter estimates—such as the average population effect size estimate M_ρ, between-context variance tau square τ^2, and sampling variance V_ρ—have been obtained, Wald-type significant tests and 95% confidence intervals (*CI*s) are then calculated for the population effect size estimate under the assumption of normality (Hedges & Vevea, 1998; Hunter & Schmidt, 2004, p. 205). Cohen (1988) developed a series of rule of thumb when talking about the magnitude of effect sizes, for the instance of M_ρ, that a small effect size has a correlation of at least .10, a medium effect size has a correlation of at least .24 but less than .37, and a large effect size has a minimum correlation of .37. Confidence intervals assess the accuracy of the estimated mean effect size and provide information on the extent to which sampling error remains in the weighted mean effect size. A 95% confidence interval excluding zero indicates that the correlation is significant different from zero. Since all the analyses were conducted based on the random-effects model, the 95% credibility intervals (*CR*) were also calculated to represent the range of true correlations across various populations under the assumption of normal distribution (Hunter & Schmidt, 2004, p. 205). The credibility intervals were calculated by taking the population effect estimate and adding to or subtracting from it the square root of the estimated population variance τ^2 multiplied by $z_{\alpha/2}$, where α is the desired probability (e.g., for a 95% interval). Credibility interval is calculated under the assumption that the value of τ^2 is known instead of estimated. When an estimated τ^2 is used in the *CR* calculation, a biased estimation could lead to the width of *CR* extremely wide and the upper bound go above unity. In addition, *CR* usually is much wider than *CI* due to that τ^2 estimate tends to be larger than population sampling variance V_ρ. Especially, τ^2 could include heterogeneity due to statistical artifacts, moderator effects, and the true difference among heterogeneous population effect sizes.

Q index. Between-study heterogeneity was tested for each of the four analysis procedures. The postulated null hypothesis for the test of *Q* statistic is that the true underlying correlation coefficient is identical for every study that is included in the meta-analysis, meaning τ^2 is not significantly different from zero. A reject of this hypothesis refers to there was a significant amount of dispersion or variability of effect sizes between studies and the extent to which effect sizes varied across studies more than would be expected due to sampling error. Heterogeneity of effect size was evaluated for statistical significance using the classical measure of heterogeneity Cochran-Mantel-Haenszel *Q* index (Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). *Q* index is calculated as the weighted sum of squared differences between individual study effects and the pooled effect across studies, with the different weighting schemes described for each analysis procedure. *Q* index is distributed as a chi-square statistic with the number of studies minus one degree of freedom. Statistically significant *Q* suggests a lack of homogeneity and there is more variability in effect sizes than expected by chance fluctuations, identifying the potential unmeasured variables as moderating the observed relationship (Cooper, 1998). These statistics also give the probability that variation in effect sizes is due to sampling error alone. However, *Q* index chi-square test has low power (smaller than 0.80) for detecting true heterogeneity when the number of studies or sample sizes is small (Filed, 2001).

I^2 index. Another commonly used index of heterogeneity is the Higgins' I^2 index, which is interpreted as the amount of heterogeneity *relative* to the total amount of variance in the observed effects or outcomes. The total amount of variance is composed of the variance in the true effects τ^2 plus sampling variance V^2_ρ (Higgins & Thompson, 2002). Judging the severity of measured heterogeneity is subjective, however, Higgins suggests some rule of thumb as a rough guide. An index between 0% and 30% is considered low heterogeneity, any value between 30% and 60% is considered moderate, and a value between 60% and 90% is considered a substantial degree of heterogeneity. However, any I^2 that falls between 75% and 100% is conventionally considered heterogeneous representing estimates drawn from multiple different populations.

Within each set of predictor-criterion study in Table 1, results from the two types of predictor-criterion reliability combinations were compared in terms of similarity and discrepancy of the estimation accuracy that each meta-analysis procedure offers. A direct comparison between the mean population effect sizes of two reliability combinations can be conducted using a Z test based on the suggestion from Raju & Brand (2003). The Z test for assessing whether two correlations are significantly different from each other can be expressed as

$$z = \frac{\hat{M}_1 - \hat{M}_2}{\sqrt{\hat{V}_1 + \hat{V}_2}}, \tag{24}$$

where M_1 and M_2 are mean population estimates from two meta-analyses; and V_1 and V_2 are the corresponding sampling variance estimates.

Statistical Programs

Reliability analysis was conducted using Microsoft Excel Macro. For the four meta-analysis procedures, Microsoft Excel spreadsheet as well as R Project for Statistical Computing the Metafor package (Viechtbauer, 2010) were utilized. Following each analysis procedure, the observed correlation and corresponding sampling variances were first prepared in Microsoft Excel especially when observed effect sizes needed to be transformed using Fisher's *z*-transformation or corrected for unreliability, and then they were applied to R Metafor for in-depth meta-analyses. When specifying the models in Metafor, heterogeneity estimator is estimated with what are suggested via the method argument, such as the DerSimonian-Laird estimator for the Hedge-Vevea procedure, and the Hunter-Schmidt estimator for the Hunter-Schmidt bare-bones procedure, the Hunter-Schmidt artifact correction procedure, and the RBNL correction procedure (Raju et al., 1991; Viechtbauer, 2005, p. 265, Viechtbauer, 2010, p. 13). Viechtbauer (2016) has stated that "there are some subtle differences between the Hunter & Schmidt method and the statistical/theoretical framework underlying the Metafor function Hunter-Schmidt estimator, so it is not an exact replication of the H&S method and the corresponding software… However, this approach still tends to work rather well." (Viechtbauer, 2016).

Summary

The complexity of the construct of Organizational Citizenship Behavior (OCB) and how the information was reported by the sample studies led to a decision of coding sample studies by variables of OCB-Overall, OCB to Individuals (OCB-I), and OCB to Organization (OCB-O), in addition to another criterion variable, Job Performance (JP). Each criterion variable was also coded separately based on the types of reliability estimates as either internal consistency or intra-rater reliability. This eventually led to a design of four sets of data for an organization of four criterion constructs and in total eight reliability combinations. A correlation between reliabilities of the predictor POS and the criterion variable was calculated for each reliability combination scenario. Four meta-analysis procedures were then implemented for each scenario. This research design facilitated the investigation of how the type of reliability and the meta-analysis procedure interact with each other and how they impact the analysis and research conclusions in practice.

CHAPTER FOUR

RESULTS

This chapter presents the quantitative analytic results for the bivariate relationship between reliability estimates of predictor and criterion variables, as well as the results from the four meta-analysis procedures for the eight reliability combinations listed in Table 1. These analyses seek to provide empirical evidences for the discrepancies due to correlated reliabilities, uncorrected or corrected effect sizes, and different analysis procedures. The first section addresses the degree of relationship between reliability estimates of predictor and criterion variables, and how these correlations compare with each other between the two types of criterion reliability for the same criterion variable. The second section addresses the meta-analyses results and the differences between the four analysis procedures for each of the four predictor-criterion studies. The third section seeks to identify the pattern emerging from the estimation differences due to the analysis procedures and the types of reliability.

Correlated Reliabilities

The 132 articles retrieved from the literature searching process rendered a total of 242 effect sizes. The overall average sample size was 292 and the sample size for each reliability combination scenario can be found in Table 2. All the studies associated with the 242 effect sizes reported information for reliability of POS. Self-rated internal consistency alpha was the only type of reliability reported for POS with estimates ranging from .7000 to .9800, an unweighted average value of .8958, and the sample size weighted average of .8932. Overall, the distribution of reliability estimates of POS, as indicated in Figure 1, formed a bell shape with a peak at .9300 and a skewness of −1.5824 as well as a kurtosis value of 2.7931. Only three studies used the 36-item full version scale developed by Eisenberger et al., (1986), and about 125 studies measured POS using a shortened version of eight or nine items (Eisenberger, Armeli, & Lynch, 1997). With respect to the four criterion variables, 200 out of the 242 studies reported scale level reliability, 30 composite reliability estimates were calculated for studies that reported information only for subscales, and the remaining studies failed to report reliability values rendering 11 missing cases.

Table 2. Basic description of reliability estimates and reliability correlation coefficients

Criterion *Y*	Number of Effect Sizes	Total Sample Sizes	POS (predictor *X*)	Criterion *Y*							Correlation between r_{XX} and r_{YY} [h]
			Average reliability [a]	Type of reliability estimates	Average reliability (*SD*) [b]	Min. [c]	Max [d]	Median [e]	Skewness [f]	Kurtosis [g]	
JP	32	9629	.8834	Alpha	.8121 (.1057)	.4900	.9320	.8261	−1.6219	3.0977	.1166**
	59	12860	.8913	Intra-rater	.8687 (.0736)	.5600	.9900	.8700	−2.0796	6.5472	.2292**
OCB-Overall	15	12312	.9093	Alpha	.8357 (.1081)	.6300	.9400	.8400	−.8442	−.4891	−.3743**
	15	3672	.8913	Intra-rater	.8940 (.0573)	.7600	.9500	.9100	−1.3226	1.1522	.0959**
OCB-I	32	11165	.8928	Alpha	.7766 (.1559)	.2759	.9200	.8200	−1.988	3.5502	.2339**
	38	7182	.9056	Intra-rater	.8329 (.1209)	.2301	.9700	.8600	−3.5161	16.7105	.0793**
OCB-O	23	8167	.8953	Alpha	.7106 (.1834)	.0769	.9300	.7400	−1.9443	5.6567	−.2512**
	28	5712	.9060	Intra-rater	.8242 (.0836)	.7000	.9400	.8050	.0535	−1.4401	.1416**

Note: Column content is as follows: [a] mean observed reliability for Perceived Organizational Support; [b] mean and standard deviation of observed reliability for criterion variable; [c] observed minimum criterion reliability; [d] observed maximum criterion reliability; [e] observed median criterion reliability; [f] skewness of observed criterion reliability; [g] kurtosis of observed criterion reliability; [h] sample size weighted Pearson correlation for Fisher's z-transformed reliability estimates; $^{**}p < 0.01$.

Figure 1. Overall distribution of the observed reliability for POS

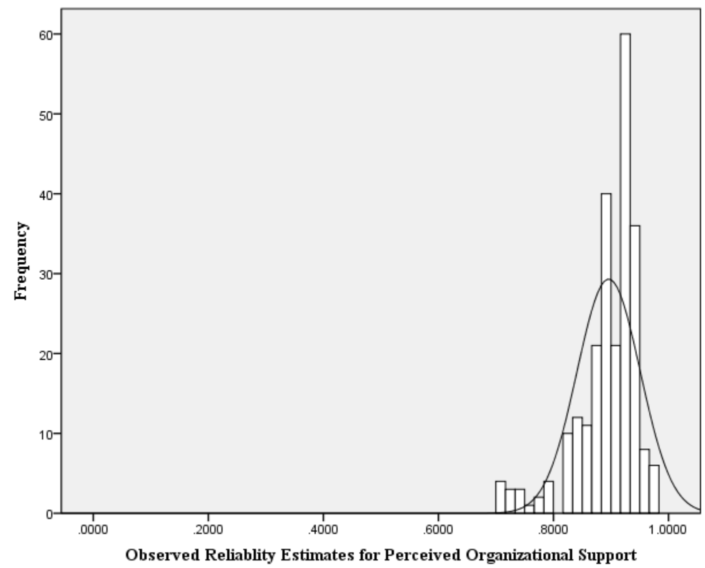


Table 2 summarizes the information on reliability estimates for the means, standard deviations, and the distributions, along with the correlation between predictor and criterion reliability coefficients across the datasets of the eight research scenarios. The averages of criterion reliabilities were quite high across the datasets. Comparatively speaking, the only exception was for the self-rated OCB-O where the average criterion reliability coefficient was slightly lower at .7106 compared to others. For the job performance variable, the average observed reliabilities were .8121 and .8687 for internal consistency and intra-rater reliability respectively. For the three organizational citizenship behavior constructs, the observed average reliabilities for internal consistency seem to be a bit higher than the average estimates of intra-rater reliability for each of the three constructs. However, it seems that higher averages of criterion reliability were observed for the combination of POS (self-rated alpha) with criterion (other-rated intra-rater) rather than the combination of POS (self-rated alpha) with criterion (self-rated alpha). As indicated by the skewness and kurtosis indices in Table 2, the distributions of these reliability estimates for the criterion variables are less than a perfect normal curve, particularly for reliability values for JP and OCB-I. Both of them have higher skewness and kurtosis indices that are far deviated from zero and three respectively, which might lead to a suppression of the correlations among reliabilities of POS and the criterion variables.

This non-normal distribution of the observed reliability values in general can affect the correlation coefficients among reliability of predictor and criterion variables. The histogram plots for the reliability estimates of predictor and criterion variables indicated a slightly left skewed unimodal shape in Figure 1 and Figure 2. A truncation in the reliability distribution for POS was manifested in Figure 1 that no value smaller than .70 was reported or observed from these studies. Therefore, a Fisher's *z*-transformation had to be applied to the observed reliabilities for both POS and its criterion variables, before a meaningful Pearson correlation can be calculated. The *z* transformation did improve the distributions of the reliabilities to a much normal distributed shape for both predictor POS and the criterion variables as a whole as displayed in Figure 3 and Figure 4.

Figure 2. Overall distribution of the observed reliability for criterion variables

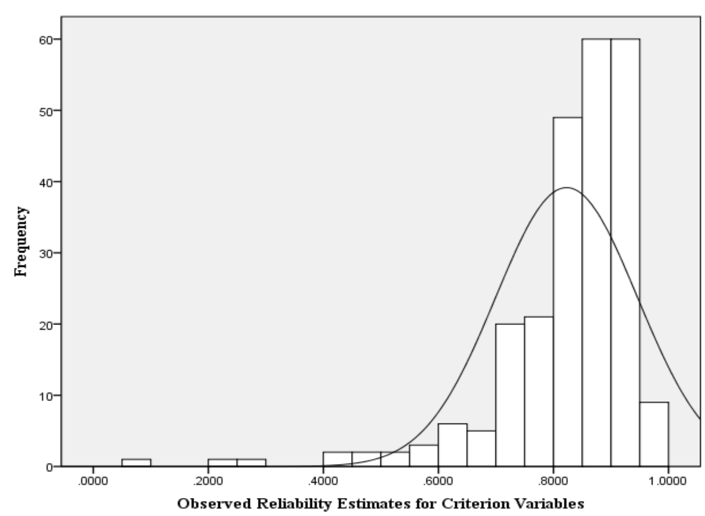


Figure 3. Overall distribution of the Fisher's *z*-transformed reliability for POS

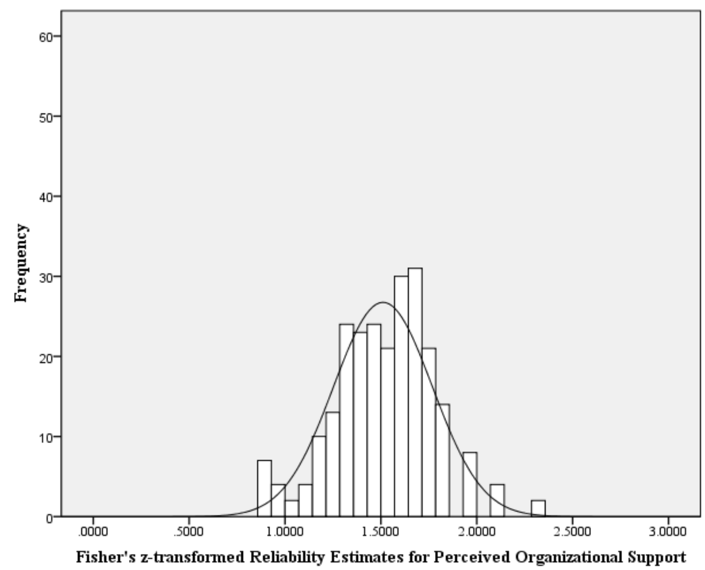


Figure 4. Overall distribution of the Fisher's *z*-transformed reliability for criterions

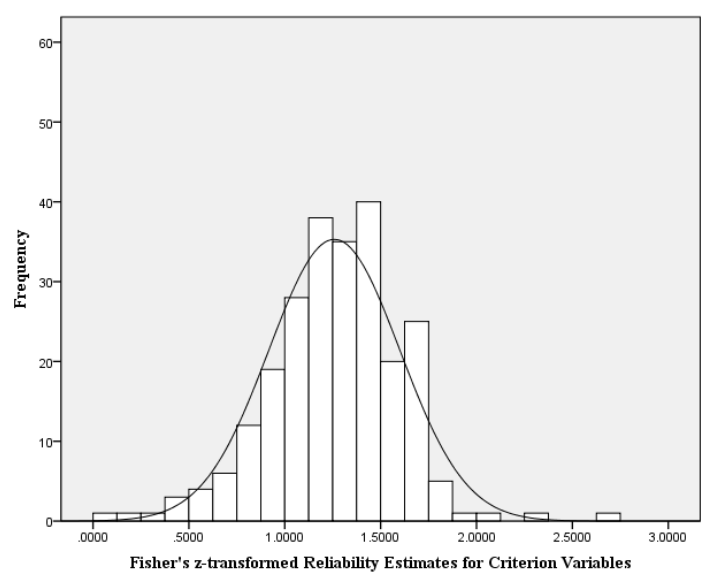


Figure 5. Scatter plot of the observed reliabilities

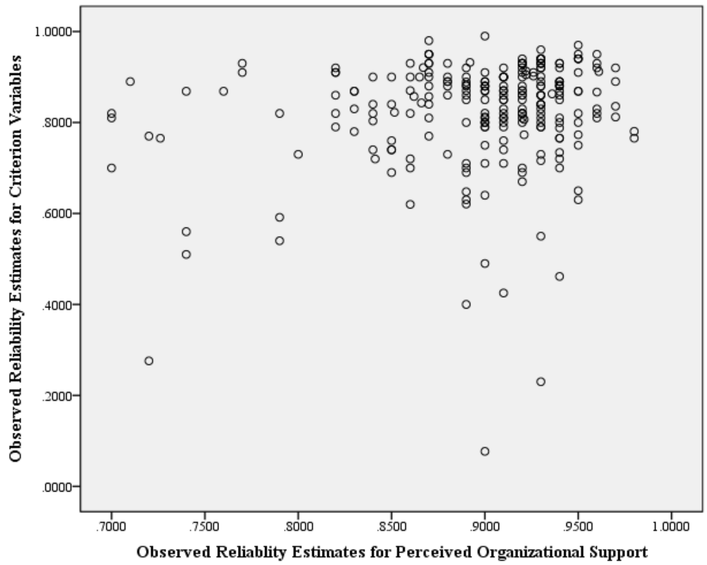
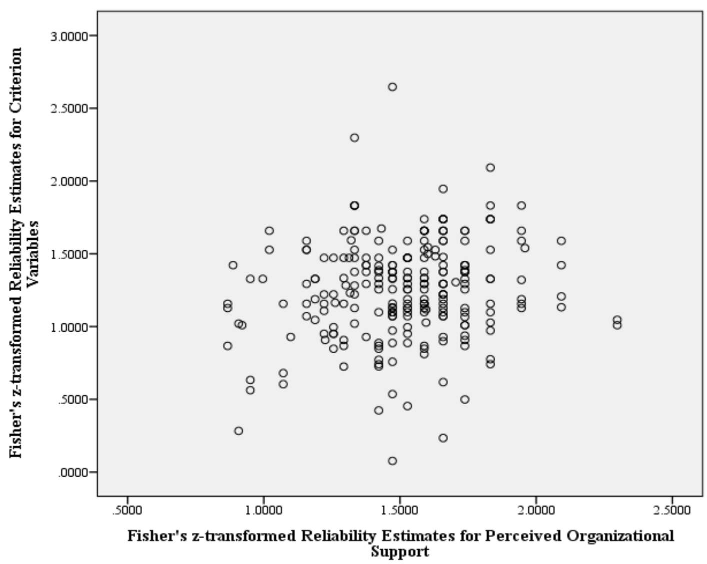


Figure 6. Scatter plot of the Fisher's *z*-transformed reliabilities



A scatter plot was adopted in order to visually detect the potential correlation between predictor and criterion reliability estimates, before a Pearson correlation coefficient was calculated. A scatter plot was conducted for the observed reliability estimates (Figure 5) and for the Fisher's *z* transformed reliability estimates (Figure 6), respectively. In general, a visual examination on both plots unfortunately did not indicate a strong linear correlation between the reliability estimates of the two variables. However, the correlation between reliability estimates of POS and its criterion variables was calculated based on the Fisher's *z*-transformed values for a sample size weighted Pearson correlation, and the results are shown in Table 2. All the eight Pearson correlations between the predictor and criterion reliability estimates were significantly different from zero based on *z* tests (Bishara & Hittner, 2012), and they differed in size and direction. Nonetheless, the largest magnitude of this reliability correlation was found between the POS internal consistency estimates and the same type of estimates for OCB-Overall, with a sample size weighted correlation of −.3743 ($p < .01$). Other substantial ones were the correlations between the POS internal consistency and JP intra-rater reliability at .2292 ($p < .01$), the correlation between the internal consistency estimates of POS and OCB-I with a value of −.2339 ($p < .01$), and the correlation between the internal consistency of POS and the same type of reliability of OCB-O with a weighted Pearson coefficient of −.2512 ($p < .01$). The magnitude of this Fisher's *z*-transformed reliability correlation manifested to a lesser extent for other combinations such as the relationship between the reliability estimates of POS internal consistency and OCB-Overall intra-rater, as well as for the correlation of POS internal consistency and OCB-I intra-rater reliability, with values of .0959 and .0793, respectively. A rough examination on these correlations did show that POS internal consistency and its criterion's internal consistency correlated to a slightly higher extent, when compared to the correlation between POS internal consistency and its criterion's intra-rater reliability.

From the distribution of criterion reliability estimates, a pattern was observed that the majority of the smaller reliability estimates happened to be the composite ones. For example, the seven low criterion reliability estimates (lower than .5) were all composite scores, with three of them from estimates of OCB-I internal consistency and the lowest value of the seven was .0769 observed for OCB-O internal consistency. When excluding the three low composite reliability scores from calculation for reliability correlation, the Pearson correlation for the Fisher's *z*-transformed POS internal consistency and OCB-I internal consistency decreased from .2339 to −.0812, not only a change in the strength but also in the direction. The magnitude of the correlation between transformed reliability of POS internal consistency and OCB-O internal consistency slightly increased from −.2512 to −.2585, after removing the smallest composite reliability value of .0769. Last but not least, it was worth noting that 12 out of the total 30 composite reliabilities were for OCB-I intra-rater estimates. On the condition of removing these 12 composite scores from this set of data, the remaining 26 observed reliability scores for the pair of POS internal consistency and OCB-I intra-rater reliability offered an increase in the Pearson correlation to .3064 from the original correlation coefficient of .0793. All these are holistically evidencing that how composite scores are calculated might have a significant impact on the correlation of reliability estimates.

Meta-Analysis Results

Three general focuses of the meta-analysis results were addressed here including weighted average, variability, and predication for the eight study scenarios identified in Table 1 and Table 2. The weighted average relates to the expected magnitude of effect size M_ρ across a large population of participants and studies. The variation V_ρ associates with the estimated true population effect size M_ρ facilitating a test of the significance of this average, and the 95% confidence interval (95% *CI*) around the average is normally provided along with M_ρ and V_ρ. The focus on variability was manifested through another set of key meta-analysis estimates including the between study variance τ^2, as well as the Q test for heterogeneity and Higgins' I^2. Lastly, the prediction relates to the 95% credibility or predict interval (95% *CR*) for M_ρ.

The Relationship between POS and JP

Four meta-analysis procedures were implemented to analyze data collected for the investigation of the relationship between perceived organizational support and one of its consequences, job performance (Kurtessis, Eisenberger, Ford, Buffardi, Steward, & Adis, 2015). The results of the meta-analyses on these data are given in Table 3.

Table 3. Results of the four meta-analysis procedures for the correlation of POS with JP by the two types of criterion reliability estimates

Criterion Y (reliability type)		No. of effect sizes	Total No. of participants	Hedges-Vevea estimations	Hunter-Schmidt Bare-Bones estimations	Hunter-Schmidt Individual Artifact Correction estimations	RBNL Individual Artifact Correction estimations
JP		32	9629				
(self-rated alpha)	M_ρ			.2169**	.1859**	.2168**	.2168**
	V_ρ			.0035	.0032	.0045	.0040
	τ^2			.0238	.0183	.0237	.0321
	95% *CR*			−.0876, .4842	−.0880, .4598	−.0948, .5283	−.1454, .5789
	95% *CI*			.1602, .2721	.1184, .2535	.1396, .2940	.1282, .3054
	Q			244.4487**	220.8425**	204.6428**	292.7216**
	Higgin's I^2			87.3184%	85.1371%	84.0026%	88.8462%
JP		59	12860				
(other-rated intra-rater)	M_ρ			.1653**	.1566**	.1824**	.1824**
	V_ρ			.0047	.0044	.0058	.0055
	τ^2			.0107	.0094	.0128	.0151
	95% *CR*			−.0388, .3561	−.0365, .3497	−.0431, .4079	−.0623, .4271
	95% *CI*			.1330, .1972	.1229, .1902	.1431, .2217	.1407, .2241
	Q			190.7647**	185.6367**	189.4439**	220.9043**
	Higgin's I^2			69.5961%	68.0718%	68.7409%	73.1895%

Note: ** $p < .01$

Estimated mean effect sizes. It is consistently observed across the four analysis procedures for both types of criterion reliability estimates that POS had a small yet statistically significant correlation ($p < .01$) in the predicted direction with JP. This significant relationship was also captured by the fact that the lower bound of 95% confidence intervals for M_ρ were above zero for both sets of data by the four analysis procedures. Estimates of M_ρ varied very slightly depending on which procedure was used. The average population effect size estimates ranged from .1859 to .2169 for the reliability combination of POS (self-rated alpha) and JP (self-rated alpha), and the same parameter for the correlation of POS (self-rated alpha) with JP (intra-rater) has estimates varying from .1566 to .1824. Specifically, for the correlation of POS (self-rated alpha) and JP (self-rated alpha), M_ρ estimates from the two artifact correction procedures were almost identical to the Hedges-Vevea analysis result with a difference of .0001. All the three estimates were greater than that from the Hunter-Schmidt bare-bones procedure by about .0300 units, and this amount did not represent a statistically significant difference. For the correlation of POS (self-rated alpha) and JP (intra-rater), the two artifact correction procedures continued providing the same estimates for M_ρ at .1824 which was slightly yet not significantly higher than that from the Hedges-Vevea's procedure by .0171 units, and the lowest value of M_ρ estimate was again found at the Hunter-Schmidt's bare-bones method as .1566. In general, the Hunter-Schmidt artifact correction and the RBNL artifact correction generated similar estimates for M_ρ. However, when reliability artifacts are not considered, the Hunter-Schmidt bare-bones method produced smaller estimates compared to the Hedges-Vevea estimates which actually were quite close to the results of the two artifact correction procedures. The Hunter-Schmidt bare-bones and the Hedges-Vevea estimates of M_ρ were slightly smaller than that of the Hunter-

Schmidt artifact correction and the RBNL artifact correction estimates. But none of the differences in M_ρ estimates among the four analysis procedures were statistically significant.

When considering how the types of criterion reliability differed between the two sets of data, the data for the correlation of POS (self-rated alpha) with JP (self-rated alpha) generally provided higher M_ρ estimates than the data for the correlation of POS (self-rated alpha) with JP (intra- rater). The largest difference occurred at the results of the Hedge-Vevea procedure by .0516 units and the smallest difference in the M_ρ estimates of the two reliability combinations came from the Hunter-Schmidt bare-bones procedure by .0293 units. The differences in the estimates of M_ρ due to the meta-analysis procedures were generally smaller compared with the differences due to the types of criterion reliability. But again none of the differences in M_ρ estimates were statistically significant.

Sampling variances. Sampling variance estimates V_ρ also varied slightly within the four procedures for the two reliability combination scenarios. In both cases, the Hunter-Schmidt bare-bones produced smaller V_ρ estimates than those from any other three procedures, and the highest estimates for V_ρ came from the Hunter-Schmidt artifact correction procedure. For example, looking closely at the combination of POS (self-rated alpha) with JP (self-rated alpha), the lowest V_ρ was at .0032 from the Hunter-Schmidt bare-bones procedure, followed by .0035 from the Hedges-Vevea procedure and .0040 from the RBNL estimates, and the highest V_ρ was at .0045 from the Hunter-Schmidt artifact correction procedure. When looking at the results from the combination of POS (self-rated alpha) with JP (intra-rater), the Hunter-Schmidt bare-bones procedure again generated the lowest estimates of V_ρ at .0044 while the Hunter-Schmidt artifact correction procedure generated the highest estimate of V_ρ at .0058. The differences in estimates of V_ρ across the no-artifact-correction procedures were slightly greater than the differences across the correction procedures in both cases of reliability combination. The V_ρ estimates for the combination of POS (self-rated alpha) with JP (self-rated alpha) were lower for each procedure than those from the comparable procedures for the combination of POS (self-rated alpha) with JP (intra-rater). Specifically, the greatest difference between the two types of combination in estimates of V_ρ occurred at the results of the RBNL procedure by .0015 units. In general, the differences in the estimates due to the meta-analysis procedures were smaller compared with the differences due to the two reliability combinations.

Credibility intervals. Lower bound of credibility interval is considered as a key indicator to infer whether validity generalizes or there is a true difference in effect sizes other than sampling errors from one population to another population, which in turn determines if the effect of situational moderators needs to be examined. A crucial difference occurs in practice particularly when the lower bound falls above zero for one analysis, meaning validity generalizes and situational moderators do not substantially exist, and falls below zero for another analysis of the same data which means validity does not generalize and situational effects need to be investigated. In such cases, there may be a sharp difference in conclusions, regardless the two credibility intervals overlap to a substantial degree. The numbers in Table 3 show that both the choice of analysis procedures and the choice of artifact correction did not significantly affect the conclusions of the study for the relationship of POS with JP. Data from the two reliability combinations of POS with JP concluded the same—that there was a chance the relationship between POS and JP did not vary across different populations regardless how the reliability of JP was estimated, because the lower bounds of the credibility intervals were below zero in every case of analysis and all the upper bounds of the credibility intervals were above zero.

However, the choice of artifact correction slightly influences the width of credibility intervals due to the differences in τ^2. For the results of the two types of criterion reliability, the credibility intervals based on data corrected for unreliability were wider than those from data without artifact correction. For example, the width of credibility intervals for POS (self-rated alpha) with JP (self-rated alpha) was .6231 and .7243 for each of the Hunter-Schmidt and the RBNL artifact correction procedures, respectively, and the widths calculated from the bare-bones procedures were .5718 and .5478 for the Hedges-Vevea and the Hunter-Schmidt bare-bones procedures. The RBNL procedure provided the largest τ^2 estimates and the Hunter-Schmidt bare-bones procedure provided the smallest τ^2 estimates. This pattern is also true for results for the correlation of POS (self-rated alpha) with JP (intra-rater).

In the same vein, the choice of criterion reliability also slightly influences the width of credibility intervals. For example, the combination of POS (self-rated alpha) with JP (intra-rater) generated narrower confidence intervals with widths ranging from .3862 to .4894 when compared to the widths of the credibility intervals for the combination of POS (self-rated alpha) with JP (self-rated alpha) which ranged from .5478 to .7243.

Degree of heterogeneity. Q tests of homogeneity and I^2 were used to determine whether the variance in population effect sizes was different from zero. The Q statistic indicates the presence of heterogeneity between effect sizes, but it does not provide information about the extent of that heterogeneity. Q tests also has low power especially when small number of studies are included in a meta-analysis. As a complement, the I^2 statistic estimates how much of the total variability in the effect size estimates (which is composed of heterogeneity and sampling variability) can be attributed to heterogeneity among the true effects (Higgins & Thompson, 2002). It is important to realize that I^2 is often estimated imprecisely, especially when the number of studies is small. As seen in Table 3, each of the *Q* tests was statistically significant ($p < .01$) with high heterogeneity coefficients I^2 coefficients ranging from 84.0026% to 88.8462% for POS (self-rated alpha) with JP (self-rated alpha) and from 68.0718% to 73.1895% for POS (self-rated alpha) with JP (intra-rater), indicating there is a high degree of true between-study heterogeneity. The heterogeneity tests with statistically significant *Q* statistics and high I^2 coefficients are presumably representing estimates drawn from multiple populations. Because there are noteworthy variances in the effect sizes for both POS (self-rated alpha) with JP (self-rated alpha) and POS (self-rated) with JP (intra-rater), a logical step forward is to examine what sample and study characteristics might best explain that variability. However, it should be noted that the measure of POS (self-rated alpha) with JP (intra-rater) contained relatively lower I^2 than the heterogeneity coefficients from POS (self-rated alpha) with JP (self-rated alpha).

The Relationship between POS and OCB-Overall Scores

The four meta-analysis procedures were implemented to analyze data for the relationship between POS and OCB-Overall, and the results are given in Table 4.

Estimated mean effect sizes. First to note, it is observed in Table 4 that POS had a positive and statistically significant relationship ($p < .01$) with another criterion variable, OCB-Overall for both reliability combination scenarios. The lower bound of 95% confidence intervals for M_ρ were above zero for both sets of data by the four analysis procedures. If the estimated correlation between POS and JP in Table 3 is considered as low based on Cohen's rule of thumb, then the estimated correlations between POS and OCB-Overall from Table 4 definitely fall into the range of moderate defined by Cohen as .24 to .37. The estimates of M_ρ varied slightly between different procedures, depending on the choice of meta-analysis procedures and the decision on adopting artifact correction. The average population estimates ranged from .2881 to .3538 for the combination of POS (self-rated alpha) with OCB-Overall (self-rated alpha), and the estimates for the combination of POS (self-rated alpha) with OCB-Overall (intra-rater) varied from .2014 to .2264. In general, estimates from the two artifact correction procedures were larger than these from the bare-bones procedures for both cases. For the combination of POS (self-rated alpha) with OCB-Overall (self-rated alpha), the Hunter-Schmidt artifact correction again provided the same estimate of M_ρ with the RBNL procedure at .3538. The Hedges-Vevea procedure provided an estimated value very close to the values from the two artifact correction procedure at .3460 which however is larger than the result from the Hunter-Schmidt bare-bones procedure by .0579 units. Yet, none of these differences were statistically significant. The similar pattern was observed for the combination of POS (self-rated alpha) with OCB-Overall (intra-rater), where the differences between the results of the four procedures were less salient in that the artifact correction procedures provided the highest estimate at .2264 which was only .0009 units different from the one from the Hedges-Vevea procedure. The lowest estimate was again produced from the Hunter-Schmidt bare-bones procedure at .2014.

When comparing the results of the two types of reliability combination, the largest difference occurred at the two artifact correction procedures by .1274 units. The smallest difference came from the Hunter-Schmidt bare-bones procedure by .0867 units between the two scenarios. The differences in the estimates of M_ρ caused by the four meta-analysis procedures were generally smaller compared to the differences between the results of the two sets reliability combinations. Data of POS (self-rated alpha) with OCB-Overall (self-rated alpha) provided slightly higher mean population effect size estimates, in contrast to the data from the combination of POS (self-rated alpha) with OCB-Overall (intra-rater). Yet, this difference was not statistically significant.

Table 4. Results of the four meta-analysis procedures for the correlation of POS with OCB-Overall by the two types of criterion reliability estimates

Criterion Y (reliability type)		No. of effect sizes	Total No. of participants	Hedges-Vevea estimations	Hunter-Schmidt Bare-Bones estimations	Hunter-Schmidt Individual Artifact Correction estimations	RBNL Individual Artifact Correction estimations
OCB-Overall (self-rated alpha)		15	12312				
	M_ρ			.3460**	.2881**	.3538**	.3538*
	V_ρ			.0018	.0015	.0021	.0012
	τ^2			.0500	.0138	.0178	.0783
	95% *CR*			−.0935, .6727	.0198, .5563	.0484, .6592	−.2850, .9926
	95% *CI*			.2360, .4473	.1499, .4262	.1964, .5112	.0261, .6815
	Q			408.1835**	216.3360**	185.3613**	1257.7741**
	I^2			96.5702%	90.2474%	89.3893%	98.4993%
OCB-Overall (other-rated intra-rater)		15	3672				
	M_ρ			.2255**	.2014**	.2264**	.2264**
	V_ρ			.0042	.0039	.0049	.0041
	τ^2			.0278	.0149	.0186	.0577
	95% *CR*			−.1093, .5144	−.0498, .4527	−.0546, .5074	−.2655, .7183
	95% *CI*			.1369, .3105	.1247, .2782	.1404, .3123	.0841, .3687
	Q			105.9691**	74.1916**	73.7548**	232.8618**
	I^2			86.7886%	79.4393%	79.3150%	93.3520%

Note: $^*p < .05$; $^{**}\, p < .01$

Sampling variances. Consistent with those findings from Table 3 for the relationship of POS and JP, estimates of V_ρ in Table 4 also varied slightly within the four procedures for the two reliability combination scenarios. In both cases, the Hunter-Schmidt artifact correction procedure produced the largest V_ρ estimates, compared with those from the other three procedures. In addition, the Hunter-Schmidt bare-bones procedure gave relatively lower V_ρ when compared with the other three, except the case of POS (self-rated alpha) with OCB-Overall (self-rated alpha). For example, looking closely at the combination of POS (self-rated alpha) with OCB-Overall (self-rated alpha), the lowest V_ρ was at .0012 from the RBNL procedure, and the Hunter-Schmidt bare-bones procedure provided the second lowest estimate of .0015. However, the pattern observed in Table 3 repeated itself here when looking at the results from the combination of POS (self-rated alpha) with OCB-Overall (intra-rater) in Table 4. The Hunter-Schmidt bare-bones procedure provided the lowest V_ρ estimates at .0039 when the highest estimate of V_ρ came from the Hunter-Schmidt artifact correction procedure with a value of .0049. The differences in estimates of V_ρ across the no-artifact-correction procedures were slightly smaller than the differences across the correction procedures, for both cases of reliability combination. Again, when the reliability of OCB-Overall was estimated through intra-rater ratings, the sampling variance V_ρ estimates were actually higher than the estimates from the data in which the reliability of OCB-Overall was estimated through self-rated internal consistency. This is the same as what was observed in Table 3 that the differences in the estimates due to the meta-analysis procedures were smaller when compared with the differences due to the reliability combinations.

Credibility intervals. As mentioned earlier, a crucial difference in conclusions occurs when the lower bound of credibility intervals falls above zero for one analysis procedure and falls below zero using another analysis procedure on the same data. Interestingly, such differences were observed from these analytical procedures for the combination of POS (self-rated alpha) with OCB-Overall (self-rated alpha). For the combination of POS (self-rated alpha) with OCB-Overall (self-rated alpha), the upper bound estimates from the four analysis procedures were all above zero; however, the lower bound of the credibility interval was a negative value of −.0935for the Hedges-Vevea procedure and a negative value of −.2850 for the RBNL procedure, yet positive lower bounds were observed for the Hunter-Schmidt bare-bones procedure at .0198 and .0484 for the Hunter-Schmidt artifact correction procedure. This indicates that practical conclusions from the Hedges-Vevea procedure and the RBNL procedure will be different from those of the two Hunter-Schmidt procedures. Situational moderators will need to be examined if one analyzes data using the Hedges-Vevea or the RBNL procedure, whereas one might choose not to explore the situational effect on population correlation if one analyzes the same data using either of the two Hunter-Schmidt methods. τ^2 estimates were higher when compared estimation values from the Hedges-Vevea procedure and the RBNL procedure with these from the two Hunter-Schmidt procedures for the combination of POS (self-rated alpha) with OCB-Overall (self-rated alpha). Variation in τ^2 estimates was a substantial contributor to the differences in the credibility intervals for the four methods. Conversely, it was more consistent across the four procedures that all the lower bound estimates were below zero and upper bound estimates were above zero for POS (self-rated alpha) with OCB-Overall (intra-rater). The data in Table 4 show that the choice of analysis procedures, the choice of artifact corrections, and the choice of how reliability of OCB-Overall was estimated influenced the conclusion about variation in the relationship between POS and OCB-Overall across different populations. Contrary to findings from Table 3, in Table 4 the credibility intervals based on data corrected for unreliability were not necessarily always wider than those from data without artifact correction. This was because the τ^2 estimates were so different across the four procedures. In Table 3, the RBNL procedure provided the largest τ^2 estimates followed by the values from the Hunter-Schmidt artifact correction, the Hedges-Vevea procedure, and then the lowest from the Hunter-Schmidt bare-bones procedure. However, in Table 4, the RBNL procedure still provided the largest estimates of τ^2 while the Hunter-Schmidt bare-bones provided the lowest τ^2 estimates, but now the Hedges-Vevea procedures provided slightly larger value of τ^2 estimates than the Hunter-Schmidt artifact correction procedure.

Degree of heterogeneity. As seen in Table 4, both reliability combination scenarios included only 15 studies. It is important to note that I^2 is often estimated imprecisely and the power of Q tests of homogeneity will also decrease when the number of studies is small. Regardless, each of the Q tests were statistically significant ($p < .01$) with high heterogeneity coefficients I^2 ranging from 89.3893% to 96.5702% for POS (self-rated alpha) with OCB-Overall (self-rated alpha), and from 79.3150% to 93.3520% for POS (self-rated alpha) with OCB-Overall (intra-rater), indicating there is a high degree of true between-study heterogeneity. Again, it is noted that the measure of POS (self-rated alpha) with OCB-Overall (intra-rater) contained relatively lower I^2 coefficients than those from the correlation of POS (self-rated alpha) with OCB-Overall (self-rated alpha).

The Relationship between POS and OCB to Individuals

Estimated mean effect sizes. The meta-analysis results for the relationship between POS and OCB-Individuals are depicted in Table 5. The results of the four procedures on the two sets of data revealed the same pattern that have been observed earlier. The two artifact correction procedures produced the same estimates of M_ρ, and the lowest estimates for M_ρ was again observed at the Hunter-Schmidt bare-bones procedure for the correlation of POS and OCB-I for both sets of reliability combination. In general, there was a statistically significant correlation between POS and OCB-I in the population, regardless of how the reliability of OCB-I was estimated or what analysis procedures was applied. For the POS (self-rated alpha) with OCB-I (self-rated alpha) scenario, POS had high and positive relationship with OCB-I with effect size estimates ranging from .2466 to .4183. When the reliability of OCB-I was estimated by intra-rater coefficient, the estimated population effect sizes from the four procedures varied from .1887 to .2157 which were considered smaller effect sizes according to Cohen's (1988) conventions. Several two-tailed *z* tests revealed that the difference in the estimates of M_ρ between the two reliability combinations were significant ($p < .05$) with *p* values lower than .03 for the two Hunter-Schmidt procedures as well as the RBNL analysis procedure. Admittedly, the Type I error was not taken into consideration here for this series of *z* tests. Overall, the estimates of M_ρ did not differ significantly between the analysis procedures, however they did vary substantially between the two types of reliability combination scenarios. The differences in the estimates of M_ρ due to the meta-analysis procedures were generally smaller compared to the differences due to the reliability combinations.

Table 5. Results of the four meta-analysis procedures for the correlation of POS with OCB-Individuals by the two types of criterion reliability estimates

Criterion Y (reliability type)		No. of effect sizes	Total No. of participants	Hedges-Vevea estimations	Hunter-Schmidt Bare-Bones estimations	Hunter-Schmidt Individual Artifact Correction estimations	RBNL Individual Artifact Correction estimations
OCB-I		32	11165				
(self-rated alpha)	M_ρ			.2466**	.3577**	.4183**	.4183**
	V_ρ			.0031	.0023	.0034	.0025
	τ^2			.1590	.0720	.0909	.0649
	95% *CR*			−.4948, .7802	−.1926, .9081	−.2001, 1.0366	−.1044, .9409
	95% *CI*			.1111, .3731	.1960, .5195	.2363, .6002	.2642, .5723
	Q			1624.7754**	1086.7665**	972.6594**	1965.8416**
	I^2			98.0920%	96.8582%	96.4317%	96.2517%
OCB-I		38	7182				
(other-rated intra-rater)	M_ρ			.1941**	.1887**	.2157**	.2157**
	V_ρ			.0054	.0050	.0065	.0033
	τ^2			.0908	.0318	.0412	.0400
	95% *CR*			−.3819, .6614	−.1671, .5446	−.1893, .6206	−.1834, .6147
	95% *CI*			.0970, .2874	.1229, .2545	.1406, .2907	.1418, .2895
	Q			658.5144**	282.5479**	280.2667**	4031.6994**
	I^2			94.3813%	86.4904%	86.3675%	92.3247%

Note: *$p < .05$; ** $p < .01$

Sampling variances. Consistent with those findings from Table 4 for the relationship of POS and OCB-Overall, estimates of V_ρ for the relationship of POS and OCB-I also varied slightly within the four procedures for the two reliability combination scenarios. In both cases, the Hunter-Schmidt artifact correction produced largest V_ρ estimates among the four analysis procedures. For example, looking closely at the combination of POS (self-rated alpha) with OCB-I (self-rated alpha), the highest V_ρ was at .0034 from the Hunter-Schmidt artifact correction procedure, and the lowest V_ρ was at .0023 from the Hunter-Schmidt bare-bones method. When looking at the results from the combination of POS (self-rated alpha) with OCB-I (intra-rater), the largest estimate was again obtained from the Hunter-Schmidt artifact correction procedure at .0065; however, the lowest estimate of V_ρ was .0033 from the RBNL artifact correction procedure. The V_ρ estimates for the combination of POS (self-rated alpha) with OCB-I (self-rated alpha) was actually slightly lower than the estimates for the combination of POS (self-rated alpha) with OCB-I (intra-rater) for all the four comparable procedures. Consistent with the pattern observed at Table 3 and Table 4, the differences in the estimates due to the meta-analysis procedures were smaller when compared with the differences due to the reliability combinations.

Credibility intervals. What has been observed previously for the τ^2 estimates was seen again at Table 5. Conclusions for the generalizability of the effect sizes were same across all four analysis procedures and the two reliability combination scenarios. For both the combination of POS (self-rated alpha) with OCB-I (self-rated alpha) and the combination of POS (self-rated alpha) with OCB-I (intra-rater), the credibility intervals from the four methods contain zero. This indicates that there is a chance that the true population effect size could be zero in certain situations or study contexts. In other words, the situational moderators do matter substantially regardless of what analysis procedure was used or how the reliability was estimated. The data in Table 5 show that the choice of analysis procedures, the choice of artifact corrections, and the choice of how OCB-I is estimated did not influence the conclusion about the variation of the correlation for POS with OCB-I caused by situational moderators in a population of studies. The credibility intervals based on data corrected for unreliability were not necessarily always wider than those from data without artifact correction. For example, the width of credibility intervals for POS (self-rated alpha) with OCB-I (self-rated) alpha was 1.0453 for the RBNL correction procedure and 1.2367 for the Hunter-Schmidt correction procedure, and the widths calculated form the bare-bones procedures were 1.2750 and 1.1007 for the Hedges-Vevea and the Hunter-Schmidt bare-bones procedures, respectively. For the correlation between POS (self-rated alpha) with OCB-I (intra-rater), the Hedges-Vevea procedure provided the widest creditability interval of 1.0433 whereas the Hunter-Schmidt bare-bones procedure generated the narrowest credibility interval of .7117 out of the four procedures. When comparing the credibility intervals of the two reliability combination scenarios, the interval for the combination of POS (self-rated alpha) with OCB-I (self-rated alpha) was wider than the one for the combination of POS (self-rated alpha) with OCB-I (intra-rater).

In addition, a suspicious overestimation for the upper bound of credibility intervals was spotted for the combination of POS (self-rated alpha) with OCBI- (self-rated alpha) using the method of Hunter-Schmidt artifact correction with a value of 1.0366 larger than 1. A correlation coefficient of unity indicates 100% predictive relationship for one construct to another, which is impossible in reality unless the two constructs are exactly the same. This overestimation will be discussed further in Chapter Five.

Degree of heterogeneity. Q tests of homogeneity and I^2 were used to determine the degree of variance in population effect sizes. As seen in Table 5, each of the Q tests was statistically significant ($p < .01$) with high heterogeneity coefficients I^2 ranging from 96.2517% to 98.0930% for POS (self-rated alpha) with OCB-I (self-rated alpha) and from 86.4904% to 94.3813% for POS (self-rated alpha) with OCB- (intra-rater), indicating the existence of a high degree of true between-study heterogeneity. This warrants the next step to examine what sample and study characteristics might best explain that variability. It should be noted that the measure of POS (self-rated alpha) with OCB-I (intra-rater) contained relatively lower I^2 than that of POS (self-rated alpha) with OCB-I (self-rated alpha).

The Relationship between POS and OCB to Organization

The meta-analysis results were organized in Table 6 for the correlation of POS with OCB-Organization.

Estimated mean effect sizes. The estimation results from the four analysis procedures for parameter M_ρ did not vary significantly from each other for either the combination of POS (self-rated alpha) with OCB-O (self-rated alpha) or the combination of POS (self-rated alpha) with OCB-O (intra-rater). The estimates for M_ρ did not differ drastically between results from the Hunter-Schmidt artifact correction procedure and the RBNL artifact correction procedure. However, when artifacts were not considered, the Hunter-Schmidt bare-bones procedure produced slightly higher estimate at .5365 when compared to the Hedges-Vevea procedure of .3913 for the combination of POS (self-rated alpha) with OCB-O (self-rated alpha); as for POS (self-rated alpha) with OCB-O (intra-rater), the Hedges-Vevea procedure produced an estimate of M_ρ at .2357 higher than the one from the Hunter-Schmidt bare-bones procedure at .2267. The uncorrected Hunter-Schmidt and Hedges-Vevea estimates of M_ρ were usually smaller than that of the Hunter-Schmidt and the RBNL artifact corrected estimates.

Table 6. Results of the four meta-analysis procedures for the correlation of POS with OCB-Organization by the two types of criterion reliability estimates

Criterion Y (reliability type)		No. of effect sizes	Total No. of participants	Hedges-Vevea estimations	Hunter-Schmidt Bare-Bones estimations	Hunter-Schmidt Individual Artifact Correction estimations	RBNL Individual Artifact Correction estimations
OCB-O (self-rated alpha)		23	8167				
	M_ρ			.3913**	.5365**	.6464**	.6464**
	V_ρ			.0032	.0016	.0024	.0019
	τ^2			.2175	.0643	.0781	.0151
	95% *CR*			−.4783, .8735	.0396, 1.0333	.0590, 1.2338	.3876, .9053
	95% *CI*			.2171, .5414	.3447, .7282	.4346, .8582	.5515, .7414
	Q			1519.2432**	1054.8364**	876.0983**	1322.6215**
	I^2			98.5519%	97.5570%	96.9761%	88.7876%
OCB-O (other-rated intra-rater)		28	5712				
	M_ρ			.2357**	.2267**	.2600**	.2600**
	V_ρ			.0050	.0045	.0061	.0036
	τ^2			.0777	.0323	.0414	.1224
	95% *CR*			−.3063, .6624	−.1344, .5877	−.1486, .6687	−.4415, .9615
	95% *CI*			.1318, .3345	.1478, .3055	.1704, .3496	.1115, .4085
	Q			444.2714**	232.1388**	221.1883**	1646.2221**
	I^2			93.9226%	87.8319%	87.2274%	97.1367%

Note: *$p < .05$; ** $p < .01$

According to Cohen's rule of thumb, a high, positive, as well as statistically significant relationship ($p < .01$) existed between POS and OCB-O with estimated population correlations ranging from .3913 to .646. However, when the reliability of OCB-O was estimated by intra-rater coefficient, the estimated population effect sizes from the four procedures varied from .2267 to .2600 which were within the medium range according to Cohen's (1988) conventions. Similar to the patterns observed earlier, the estimates of M_{ρ} differed slightly between different procedures, but differed more significantly between the two types of reliability combinations. When comparing the results of the two types of reliability combinations, the largest and statistically significant differences ($p < .01$) in M_{ρ} estimates occurred at the results from the two artifact correction procedures by .3864 units, and the smallest raw differences between the two reliability combinations came from results of the Hedges-Vevea procedure by .1556 units which indicated a not statistically significant difference. The differences in the estimates of M_{ρ} due to the meta-analysis procedures were generally smaller compared with the differences due to the reliability combinations.

Sampling variances. Estimates of V_{ρ} also varied slightly within the four procedures for the two reliability combination scenarios for the correlation of POS with OCB-O. For the combination of POS (self-rated alpha) with OCB-O (self-rated alpha), the Hunter-Schmidt bare-bones procedure produced the smallest V_{ρ} estimates at .0032 when compared to those from the other three procedures, however the highest estimates for V_{ρ} of .0061 came from the Hunter-Schmidt artifact correction procedure for the combination of POS (self-rated alpha) with OCB-O (intra-rater). The difference in estimates of V_{ρ} across the no-artifact-correction procedures were .0016 which was greater than the differences of .0005 across the correction procedures for the combination of POS (self-rated alpha) with OCB-O (self-rated alpha). However, the difference of .0005 between the two bare-bones analysis results was smaller than the difference of .0025 between the results from the two artifact correction procedures. In general, the V_ρ estimates for the combination of POS (self-rated alpha) with OCB-O (self-rated alpha) were again higher than the estimates for the combination of POS (self-rated alpha) with OCB (intra-rater) for all the four comparable procedures.

Credibility intervals. Practical conclusions about the presence or absence of situational moderators varied greatly across the analysis results for the two types of reliability combinations. Data from the two reliability combinations of POS with OCB-O concluded differently on whether or not the relationship between POS and OCB-O varied across different populations. For the combination of POS (self-rated alpha) with OCB-O (self-rated alpha), the upper bound estimates from the four analysis procedures were all above zero; however, the lower bound of the credibility interval was a negative value of −.4783 for the Hedges-Vevea procedure, and positive values for the rest of the three analysis procedures. This indicated that practical conclusions from the Hedges-Vevea procedure could be opposite to those from the other three procedures, such that the estimated population effect size cannot be affected substantially by some situational factors. In addition, both the upper bounds of credibility intervals from the analysis of the Hunter-Schmidt bare-bones procedure and the Hunter-Schmidt artifact correction procedure exceeded unity which has a problematic meaning in practice. However, the conclusions for POS (self-rated alpha) with OCB-O (intra-rater) were more consistent across the four procedures since all the lower bound estimates were below zero and upper bound estimates were above zero. This indicated that the population effect sizes of POS (self-rated alpha) with OCB-O (intra-rater) can be substantially impacted by a situational moderator. The data in Table 6 show that the choice of analysis procedures, the choice of artifact corrections, and the choice of how reliability of OCB-O was estimated influenced the conclusion about the variation for effect sizes of POS with OCB-O in a population of studies. When compared with the credibility intervals of the two reliability combination scenarios, the widths of credibility intervals for POS (self-rated alpha) with OCB-O (self-rated alpha) were wider in comparison with the estimated widths of credibility intervals for POS (self-rated alpha) with OCB-O (intra-rater), except for the results from the RBNL procedure.

Degree of heterogeneity. Table 6 displays each of the *Q* tests as statistically significant ($p < .01$) with high heterogeneity coefficients I^2 ranging from 88.7876% to 98.5519% for POS (self-rated alpha) with OCB-O (self-rated alpha), and from 87.2274% to 97.1367% for POS (self-rated) with OCB-O (intra-rater). A high degree of true between-study heterogeneity therefore was indicated and the next step is to examine what sample and study characteristics would be able to explain this high heterogeneity. Consistent with what has been seen from Table 3 to Table 5, the measure of POS (self-rated alpha) with OCB-O (self-rated alpha) contained relatively higher I^2 than these from POS (self-rated alpha) with OCB-O (intra-rater) for each of the four analysis procedures. An exception occurred at Table 6: The RBNL generated higher degree of heterogeneity estimates of coefficients I^2 for the combination of POS (self-rated alpha) with OCB-O (intra-rater).

Inter-correlated Reliabilities and Meta-Analysis Results

In order to holistically explore how the meta-analysis results differ due to the choice of correction, the choice of analysis procedure, and the choice of reliability estimates, the key statistical indices from the previous two sections were organized together in Table 7 for each of the eight research scenarios. In addition, a method of visual display was adopted to assist this investigation. Two key meta-analysis estimators of M_ρ and V_ρ were presented in Figure 7 and Figure 8 separately for the correlation of POS with JP, and the results for the correlation between POS and the three OCB constructs were displayed in Figure 9 and Figure 10.

Table 7. Summary of the meta-analysis and the inter-correlation of reliabilities

Criterion variable/ Type of reliability	No. of effect sizes	Hedges-Vevea M_ρ (V_ρ)	Hunter-Schmidt Bare-Bones M_ρ (V_ρ)	Hunter-Schmidt Individual Artifact Correction M_ρ (V_ρ)	RBNL Individual Artifact Correction M_ρ (V_ρ)	Correlation of r_{XX} with r_{YY}
JP						
Self-rated Alpha	32	.2169 (.0035)	.1859 (.0032)	.2168 (.0045)	.2168 (.0040)	.1166
Other-rated Intra-rater	59	.1653 (.0047)	.1566 (.0044)	.1824 (.0058)	.1824 (.0055)	.2292
OCB-Overall						
Self-rated Alpha	15	.3460 (.0018)	.2881 (.0015)	.3538 (.0021)	.3538 (.0012)	−.3743
Other-rated Intra-rater	15	.2255 (.0042)	.2014 (.0039)	.2264 (.0049)	.2264 (.0041)	.0959
OCB-I						
Self-rated Alpha	32	.2466 (.0031)	.3577 (.0023)	.4183 (.0034)	.4183 (.0025)	.2339
Other-rated Intra-rater	38	.1941 (.0054)	.1887 (.0050)	.2157 (.0065)	.2157 (.0033)	.0793
OCB-O						
Self-rated Alpha	23	.3913 (.0032)	.5365 (.0016)	.6464 (.0024)	.6464 (.0019)	−.2512
Other-rated Intra-rater	28	.2357 (.0050)	.2267 (.0045)	.2600 (.0061)	.2600 (.0036)	.1416

Figure 7. The results of estimated M_ρ for the correlation of Perceived Organizational Support and Job Performance

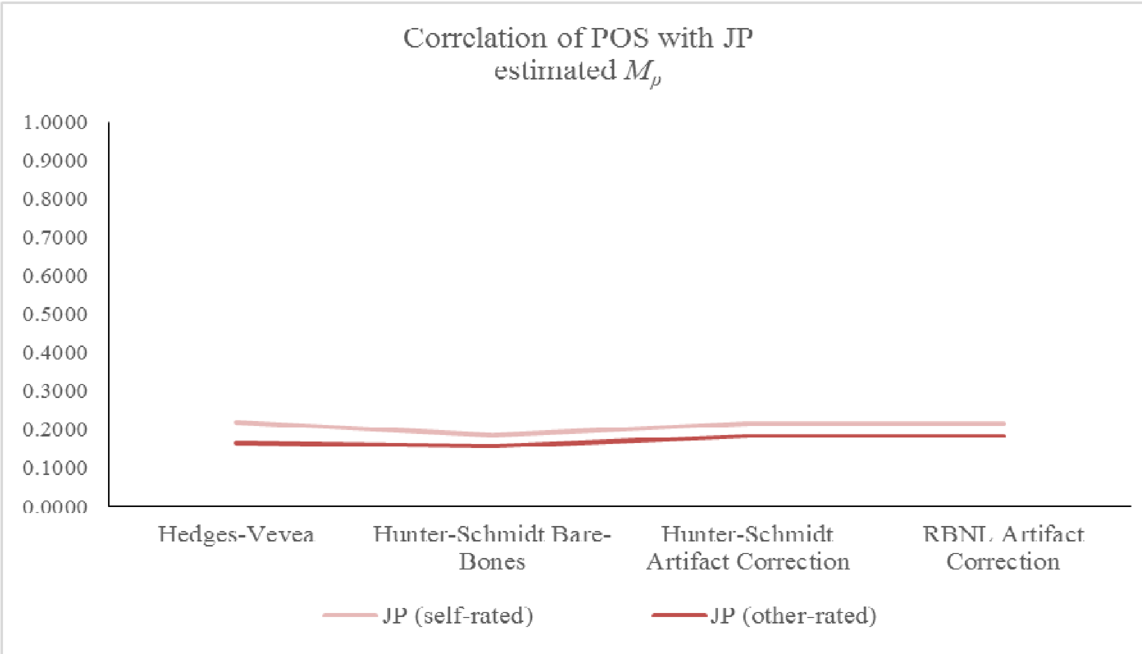


Figure 8. The results of estimated V_ρ for the correlation of Perceived Organizational Support and Job Performance

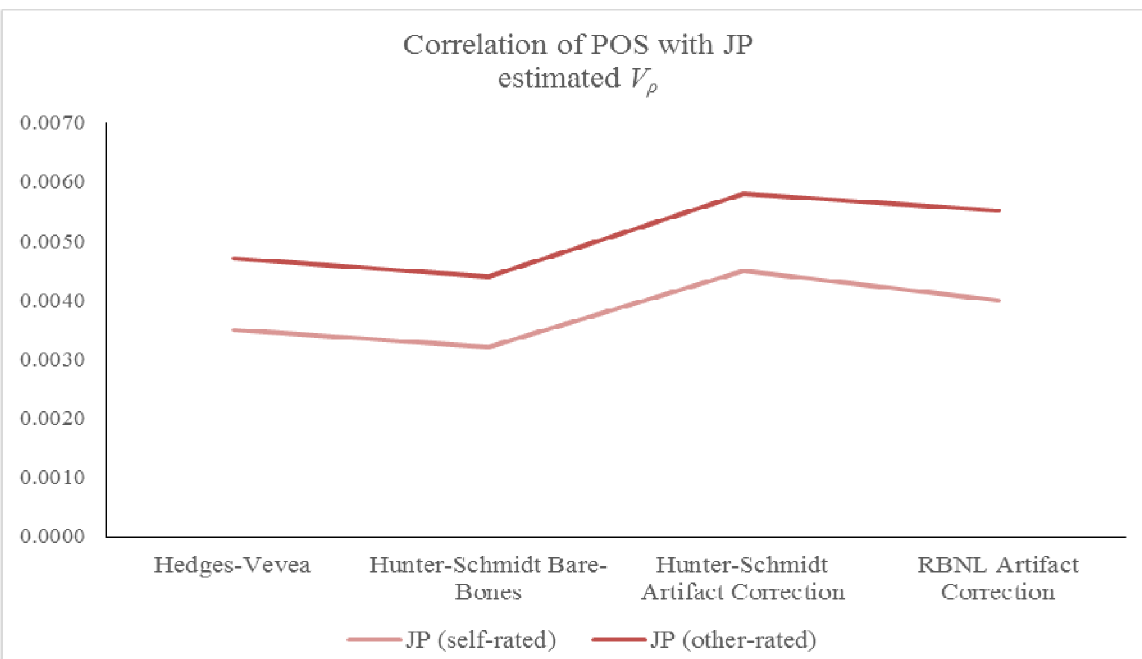


Figure 9. The results of estimated M_ρ for the correlation of Perceived Organizational Support and Organizational Citizenship Behavior Overall, OCB to Individuals, and OCB to Organization

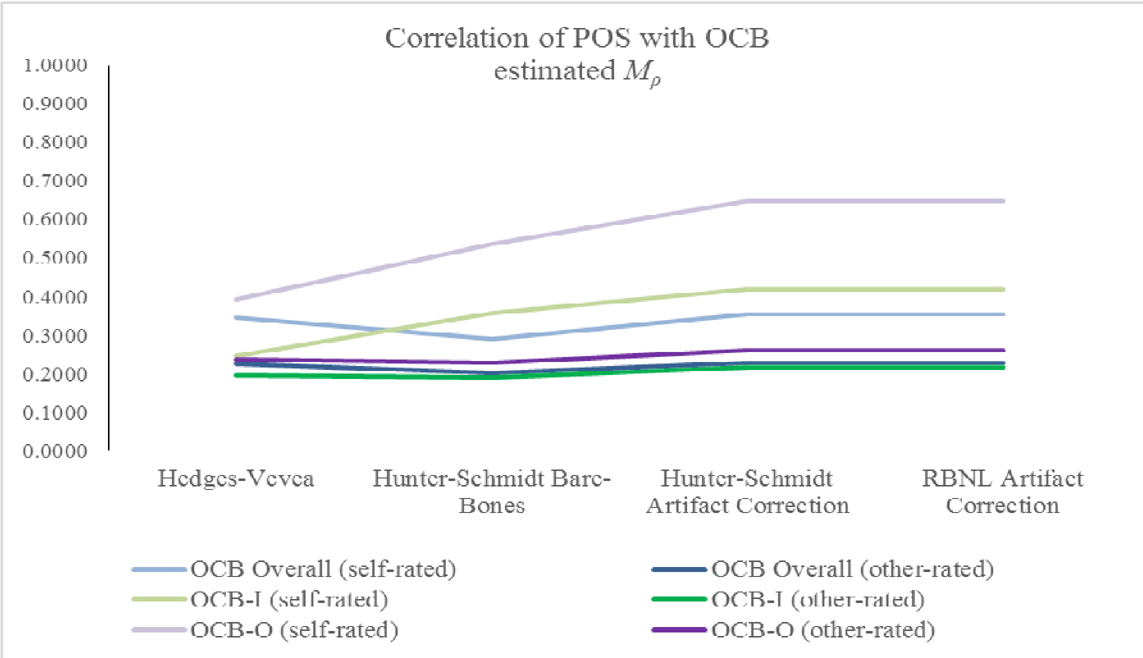
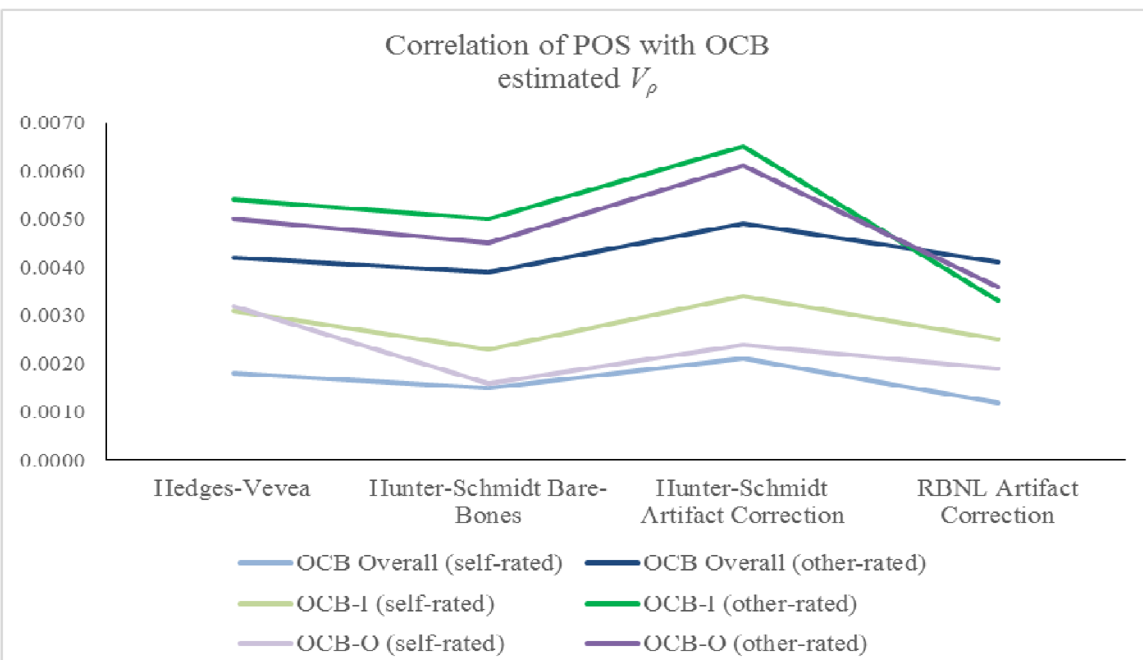


Figure 10. The results of estimated V_ρ for the correlation of Perceived Organizational Support and Organizational Citizenship Behavior Overall, OCB to Individuals, and OCB to Organization



With regard of the meta-analysis methodology, the two artifact correction procedures generated identical M_ρ estimates to four decimal places for each set of studies, and their M_ρ and V_ρ estimates were higher than those from the two bare-bones procedures. For the first two sets studying the relationships of POS with JP and POS with OCB-Overall, the differences in M_ρ between the Hedge-Vevea procedure and the two artifact correction procedure were very trivial, whereas the Hunter-Schmidt bare-bones procedure produced lower M_ρ estimates than the other three procedures. Within either of the two POS-Criterion correlations, when comparing the results between the two reliability combinations, the differences in M_ρ were very salient that the estimates from the combination of POS (self-rated alpha) with criterion variable (self-rated alpha) were higher than those from the combination of POS (self-rated alpha) with criterion variable (other-rated intra-rater), regardless of the analysis procedures or the choice of artifact correction. When looking at the estimated sampling variances V_ρ, the Hunter-Schmidt artifact correction procedure always provided the largest estimates for V_ρ among the four procedures and the lowest estimates of V_ρ were consistently generated from the Hunter-Schmidt bare-bones procedure. The studies that involved criterion variable with other-rated intra-rater reliability generated higher estimates of V_ρ when compared to the V_ρ estimates from these scenarios when criterion reliability was estimated through self-rated alpha.

For the relationship of POS with OCB-I as well as the relationship of POS with OCB-O, the two artifact correction procedures generated relatively larger estimates of M_ρ followed by those from the Hunter-Schmidt bare-bones procedure. However, the lowest values were produced by the Hedges-Vevea procedure when looking at the combination of POS (self-rated alpha) with criterion variables (self-rated alpha). When criterion reliability was estimated by other-rater intra-rater coefficient, the lowest estimates of M_ρ were produced by the Hunter-Schmidt bare-bones procedure. This pattern was different from what we have seen in the results for the correlation of POS with JP and for the correlation of POS with OCB-Overall. When looking at the estimates of V_ρ, the highest estimates were observed at the Hunter-Schmidt artifact correction procedure across the studies with the exception that for the scenario of POS (self-rated alpha) with OCB-O (self-rated alpha) the Hedge-Vevea procedure provided the highest value of V_ρ.

Summary

This chapter presents the results for the correlation analysis between reliability estimates of predictor POS and its criterion variables, as well as the meta-analysis results for the four pairs of correlations by the four different analysis procedures and the two types of reliability combinations. In general, there was a statistically significant correlation between the reliability estimates of POS and its criterion variables, although the magnitude of these correlations were moderate to small. In terms of the meta-analysis results, it was consistently observed that perceived organizational support substantially correlated with job performance and with the three constructs of organizational citizenship behavior. The choice of artifact correction, analysis procedure, and the types of reliability estimates did not significantly impact the existence of these correlations, but did impact the magnitude of these correlations and the degree of variation of the correlation in a population of studies. Major discrepancies and conflicts in practical conclusions occurred due to the differences in estimates of τ^2 for between-study variance and associated credibility interval. Certain analysis procedures might have provided overestimated upper bounds for this interval, indicating a correlation coefficient greater than unity.

CHAPTER FIVE

DISCUSSION

This chapter summarizes the results of the current study and their practical implications. The study limitations as well as the directions for future research are also discussed.

Summary of the Study

Recent reviews on meta-analysis practice have shown that the individual artifact correction approach actually has been adopted more often in field research than the artifact-distribution-based correction approach, even though a lot methodological efforts have been invested in the latter one. In fact, individual correction functionally and practically outperforms artifact-distribution-based correction methods, which linearly encourages the use of individual correction methods particularly for correcting measurement errors. The Hunter-Schmidt individual artifact correction procedure has been widely used in the Industrial and Organizational field; however, it does not specifically address the condition of correlated reliability. Raju et al.'s meta-analytic approach, on the other hand, performs quite well in the situation when there are a non-zero correlations among true validity and artifacts, and the equations for sampling variance calculation include variance components to account for the correlations between statistical artifacts (Raju et al., 1991; Raju et al., 1998). Empirically, several comparison studies adopted Monte Carlo simulations which were used to assess the robustness of the two meta-analysis procedures under various violations for pairwise artifact independence (Le & Schmidt, 2006; Mendoza & Reinhardt 1991; Raju et al., 1998). However, most these focuses were on inter-correlation between the true score and the error score, rather than the independency between reliabilities. Additionally, these studies are often very limited by their dependence of the Monte Carlo framework. That is, most assessments for the accuracy and reasonableness of conclusions from the analyses have been carried out in environments where assumptions are literally met, conditions are deliberately designed, and data is neatly simulated. None of these studies used real-world data to examine the effect of correlated reliabilities on the corrected correlation effect sizes. There is a clear gap in our knowledge of the magnitude of reliability correlations and the effect that correlated reliabilities have on the estimates of corrected mean population correlation. Based on this understanding, the current work fills in this gap by using real-world data to compare the meta-analysis results of the individual artifact correction procedures with the results of the non-correction procedures, under the condition of correlated reliabilities. This is also a good opportunity to use real data to look into the necessity of artifact correction in field study.

The current study meta-analyzed 244 effect sizes collected from the published and unpublished field research regarding the correlation of perceived organizational support and its criterion variables as an example. The four most popular meta-analysis procedures, the Hedges-Vevea procedure, the Hunter-Schmidt bare-bones procedure, the Hunter-Schmidt individual artifact correction procedure, and the RBNL individual artifact procedure by Raju et al. (1991) were applied to examine how correlated reliabilities impact on the meta-analytic results of these realistic data.

Findings

The reliability correlation was calculated based on Fisher's *z*-transformed reliability estimates and the significance of this correlation was confirmed using a *z* test. In general, there were statistically significant correlations ranging from .0793 to .3742 between the reliability of POS and the reliability of its criterion variables. Therefore, when artifact correction meta-analysis is conducted, the assumption of pairwise independence between reliabilities that most analytical procedures require cannot be carelessly assumed without examination.

Table 7 shows that estimated correlation effect sizes M_ρ varied from .1566 to .2169 in the first set of studies for the correlation of POS and JP, from .2014 to .3538 in the second set for the correlation of POS and OCB-Overall, from .1887 to .4183 in the third set for the correlation of POS and OCB-Individuals, and from .2267 to .6464 in the fourth set for the correlation of POS and OCB-Organization. High heterogeneity in the estimated population correlations was found for each set of studies regardless of the choice of analysis procedures, which in turn should lead to a further exploration for the effect of situational moderators.

When comparing the two artifact correction procedures, the Hunter-Schmidt individual artifact correction and the RBNL artifact correction, no differences were found for the estimates of M_ρ. Higher V_ρ estimate was actually always observed from the analysis of the Hunter-Schmidt artifact correction. When comparing the two non-artifact-correction methods, the Hedges-Vevea procedure and the Hunter-Schmidt bare-bones procedure, minimal differences were found between the estimates of M_ρ from the two methods, whereas higher V_ρ was observed from the Hedges-Vevea analysis procedure. Comparing across the four analysis procedures, no statistically significant differences were noted for the estimates of M_ρ, yet it did seem like the

Hunter-Schmidt artifact correction procedure always provided highest value for V_ρ and the Hunter-Schmidt bare-bones procedure consistently provided the lowest value for V_ρ. It should be noted that to date, there is no method available for the significant test for the differences of $V_{\rho.}$

When comparing the meta-analysis results between the two forms of criterion reliability, the estimates of M_ρ from the combination of POS (self-rated alpha) with criterion variable (self-rated alpha) were higher, and sometimes significantly higher, than those from the combination of POS (self-rated alpha) with criterion variable (other-rated intra-rater). The estimates of V_ρ were greater for the studies that involved the criterion variable with other-rated intra-rater reliability than those where criterion reliability was estimated through self-rated alpha.

Implications

Over the history of methodological development for psychometric meta-analysis, correlations between reliability, range restriction, and true effect sizes have been extensively studied; however, few research had depicted and studied the situation of correlated reliability. Köhler et al.'s (2015) was the first that quantitatively confirmed the existence of correlated reliability with the real-world data. Additionally, Raju et al.'s Monte Carlo study (1998) was the first of few that deliberately simulated data with predefined conditions to mimic the reality in which a pairwise independence assumption was violated to various degrees. One of the conditions they predefined for inter-correlations among artifacts was to set the magnitude of the correlation between predictor reliability and outcome reliability as .00, .25, and .35, corresponding to zero correlation, low correlation, and medium correlation respectively. They then examined the meta-analysis results and compared them under each of the three simulation conditions. Retrospectively, these specifications for their simulated data were consistent with what has been observed in the current study concerning the magnitude of the correlation between reliability estimates ranging from the lowest .0793 to the highest .3742 (Table 7). All of them were statistically significant with very small *p* values, even though it was likely that the significance was due to the large sample sizes included in the studies. Therefore, it is hardly acceptable to assume without any initial examination that there is an independent relationship between reliabilities when conduct artifact correction meta-analysis.

Raju et al. (1998) investigated how the inter-correlations among three artifacts as well as the true validity ρ impact the accuracy of meta-analytical estimation. The three artifacts that were taken into consideration included criterion unreliability, predictor unreliability, and range restriction. In addition, they analyzed the simulated data with two validity generalization models, the Hunter-Schmidt artifact correction model and the RBNL artifact correction model. In their simulation, one of the inter-correlation conditions was to set the correlation between ρ_{XX} and ρ_{YY} ranging from low negative −.2125 to high positive .4130. This simulation setup is comparable to what we have observed through the real-world data analyzed in the present study—that a correlation between ρ_{XX} and ρ_{YY} ranged from −.3743 to .2339.

Through the simulated data, they found out that the population average M_ρ was estimated similarly between the two models. Under the condition that there is non-zero correlation between artifacts and true effect sizes, the Hunter-Schmidt procedure tended to produce an overestimated V_ρ which is larger than the true V_ρ, whereas the RBNL procedure would provide a slightly lower value of V_ρ yet closer to the true value. These observations were consistent with the findings in the current study. In this study, there was no difference found between the Hunter-Schmidt individual correction procedure and the RBNL procedure for the estimates of M_ρ, however the sample variance estimates V_ρ and between-study variances τ^2 from the Hunter-Schmidt procedure were always larger than those from the RBNL procedure. The Hunter-Schmidt procedure was more likely to expose itself to a higher risk of overestimation, which in turn can easily cause an upper bound estimate of confidence intervals and credibility intervals over unity—as what was seen in Table 5 and Table 6. However, as was concluded by Raju et al. (1998), the violation of independence among artifacts has minimal effect on meta-analysis results unless artifacts are correlated with true correlations. The present study confirms that when range restriction was not considered, the differences in the estimates of M_ρ and V_ρ were trivial between the two procedures, and it seems the type of the criterion reliability or the degree of the correlation of reliabilities did not significantly impact how the two artifact correction procedures differed from each other.

Hall and Brannick (2002) also adopted a Monte Carlo framework to compare the Hedges-Vevea (1998) procedure with the Hunter-Schmidt (1990) procedure. They examined the two procedures under the situations in which there were no artifacts other than sampling error as well as the situations in which artifacts attenuated correlations due to criterion unreliability and range restriction on predictor variable. When sampling error was the only one considered, both bare-bones procedures generated comparable estimates for M_ρ. However, the Hunter-Schmidt procedure would produce less biased M_ρ than the Hedge-Vevea procedure. The estimates of V_ρ from both procedures became increasingly inaccurate as the true variance of effect sizes in the population increased. But in general, the Hunter-Schmidt estimates were more closely aligned with the true V_ρ than were those from the Hedges-Vevea procedure. They concluded that the Hedges-Vevea procedure tended to overestimate the true M_ρ, while the Hunter-Schmidt procedure produced underestimates for true V_ρ. However, the differences in estimation results by the two procedures are usually quite small. These findings were also confirmed by Field's simulation study (2001) when only sampling errors was considered. Looking at the results from the two bare-bones procedures in the present study, they were harmonically in agreement with those from the previous Monte Carlo research. The Hedges-Vevea procedure did produce higher estimates for both M_ρ and V_ρ than the Hunter-Schmidt procedure for almost all the studies, however the differences between the two procedures were trivial. The only exception was observed at the estimates of M_ρ for the combination of POS (self-rated alpha) with OCB-I (self-rated alpha) as well as the combination of POS (self-rated alpha) with OCB-O (self-rated alpha), where the Hedges-Vevea procedure actually provided lower values of M_ρ than those from the Hunter-Schmidt bare-bones procedure, yet the differences were not statistically significant. Initial examination on outliers or influential cases did not find any potential factors that might be able to explain this exception.

In the situations where artifacts (i.e., criterion unreliability and range restriction on predictor variable) were assumed, Hall and Brannick (2002) found that the uncorrected Hedges-Vevea procedure produced estimates of M_ρ and V_ρ that were about half the size of the true population values. In contrast, the Hunter-Schmidt individual artifact correction procedure produced fairly accurate estimates close to the true parameters. When looking at the results of the present studies, the uncorrected Hedge-Vevea procedure did provide lower estimates for M_ρ when compared to the Hunter-Schmidt individual artifact correction procedure. But the differences for M_ρ between the two approaches were very insignificant (i.e., .0009 and .0171) for those studies involved the correlation of POS with JP and the correlation of POS with OCB-Overall. The larger differences (i.e., .1717 and .2551) were found from the analyses for the correlation of POS (self-rated alpha) with OCB-I (self-rated alpha) as well as the correlation of POS (self-rated alpha) with OCB-O (self-rated alpha). Yet again in practical terms none of these differences were statistically significant, which was contrary to the "half the size differences" mentioned earlier by Hall and Brannick (2002). In terms of V_ρ estimates, it was interesting to see that the differences between the uncorrected Hedges-Vevea procedure and the corrected Hunter-Schmidt procedure in the present study were so small as to be negligible (i.e., the largest difference of .0011). The size of the difference in M_ρ as well as in V_ρ between the two analysis procedures in the current data were not quite aligned with what was concluded by Hall and Brannick (2002). It should be noted that Hall and Brannick's study (2002) did not factor in the situation when the pairwise independence assumption is not tenable, therefore their conclusions did not take into account of correlated reliabilities. But from the current data analysis, it seems that neither the type of reliability nor the magnitude of the inter-correlation of reliabilities alterated the conclusions reached by the previous research.

Köhler et al. (2015) pointed out that it is still unknown whether the correlated reliabilities and their interdependence actually leads to a substantial bias in meta-analytical results, and they stated that the actual effect of correlated reliabilities on the estimates of M_ρ and V_ρ in artifact correction meta-analysis is unlikely to be simple. The present study attended to address this topic and the findings could provide some insight. Although the differences between the analysis procedures were not striking and were aligned with what we have learned from the past Monte Carlo studies, the differences between the results of the two types of reliability combinations for each POS-Criterion correlation were definitely noticeable and deserved extra research attention. Regardless of the degree of reliability correlation, the analysis procedures, and the choice of artifact correction, the combination of POS (self-rated alpha) with criterion (self-rated alpha) always provided higher M_ρ and lower V_ρ estimates than the combination of POS (self-rated alpha) with criterion (other-rated intra-rater). A closer look was paid to the observed reliability values for predictor and criterion variables in Table 3. It seems that the observed average reliabilities were higher for the combination of POS (self-rated alpha) with criterion (other-rated intra-rater) than the corresponding values from the combination of POS (self-rated alpha) with criterion (self-rated alpha). When considering artifact correction, the smaller reliability estimates for the combination of POS (self-rated alpha) with criterion (self-rated alpha) were very likely inducing an overestimation for $M_{\rho,}$ and a smaller estimation for V_ρ. This might have explained why the statistically significant differences in M_ρ estimates between the two types of reliability combinations were more commonly seen in the current analyses. When the information was provided by the same person for both the predictor and criterion variable, it is likely that the correlation between the two variables would be naturally higher than the correlation from the case when the information for the two variables were provided by different parties. In the first case, some specific factor errors (i.e., all the front-line employees are under the stress of getting laid off) can be counted to the true scores for both variables, which potentially lead to a higher covariance or correlation between predicator and outcome. Holding the observed score constant, if the specific errors were counted towards the true score, this will leave a lower amount of sampling errors and variations. In personnel psychology, covariation between the two variables can be caused by the effect of situational variables such as the setting in which data is collected, the nature of personnel assessment (i.e., self-assessed or other-assessed), and some drastic organizational change. On the other hand, when the information was collected from two parties, this kind of specific error can be potentially avoided, which might lead to less covariation between variables and high variations in the relationship between them.

In the current study, the inflations for the correlation estimates and the deflations for the sampling variances seemed more salient for the combinations of predictor (self-rated alpha) with criterion (self-rated alpha), compared to the combinations of predictor (self-rated alpha) with criterion (other-rated intra-rater). It could be that how the reliability was estimated does substantially change our findings or conclusions, as opposed to the degree of correlation between reliabilities. Schmidt, Viswesvaran, and Ones (2000) insisted that interrater reliability should be used to correct an observed validity coefficient. Because it is the only reliability coefficient that reflects the effects of all four sources of measurement error that are in play, rater leniency effects, halo effects (the rater by rate interactions), random response error, and transient error. However, Murphy and DeShon (2000) preferred intra-rater reliability estimates of performance ratings such as coefficient alpha. They considered self-reported ratings, when adequately conceptualized and measured, as the most construct-valid in terms of minimizing criterion deficiency and criterion contamination. These two opposite opinions, as an example, indicated the existence of the contradiction about what kind of reliability estimates need to be used when considering artifact correction. However, it indeed is already known that reliability is sensitive to sampling characters, and it can be influenced by the purpose of measurement (e.g., managerial selection vs. developmental feedback), testing condition (e.g., noise, equipment, and time of day), and the data sources (self, other people) (Brannick & Zhang, 2013). The current research does agree that we should shift our attention to the fundamental research in reliability estimates and assess the meta-analytical differences caused by using different methods to estimate reliability for the same construct.

From the analyses on the four sets of studies and eight reliability combinations, the conclusions from the previous Monte Carlo studies can be roughly confirmed with the current research findings, particularly when looking at the differences due to various analysis procedures and the differences due to the choice of artifact correction. While the Hunter-Schmidt and the RBNL artifact correction procedures did not produce many meaningful differences, the non-correction Hedges-Vevea procedure did not differ significantly compared to the two procedures. Admittedly, the Hunter-Schmidt bare-bones procedure was the one offering lowest values for estimates of M_ρ and V_ρ, with exceptions for a few cases, such as the analyses for the correlation of POS (self-rated alpha) with OCB-I (self-rated alpha) as well as the correlation of POS (self-rated alpha) with OCB-O (self-rated alpha). The claimed primary advantage of psychometric meta-analysis (PMA) was not distinctly observed in the current research—namely, that PMA did not permit a more accurate estimate of population correlations by correcting for the statistical artifacts of measurement error. The type of reliability estimates, on the other hand, seemed to matter the most when conducting meta-analysis, regardless of whether or not artifact correction was considered. It is advocated that research attention should be shifted onto how to estimate reliability appropriately and to what extent the reliability estimates differ when various estimation methods are used. We should reflect on the research needs in reality for further development on artifact correction methods, and definitely put a pause on piling one technical refinement on top of another. Excessive refinements sometimes only lead to an increasing complexity in correction procedures and equations yet not meaningful improvement for estimating population parameters.

The claimed advantage of artifact correction was not distinctly noticeable in the current research especially when comparing the results from the Hedges-Vevea method and those from the artifact correction methods. Methodologically when estimating population parameters such as M_ρ, V_ρ, and τ^2, the differences between the no-correction Hedges-Vevea method and the other methods involve data transformation for observed effect size and the weighting scheme for calculation, especially the weighting scheme for Hedges-Vevea's is a much more complex one for a random-effect model. These calculation complexities actually make the Hedges-Vevea method a strong and robust meta-analysis method. In fact, a comparison study using Monte-Carlo simulation by Mendoza & Reinhardt (1991) concluded that the Hedge's approach performed fairly accurately when the selection ratio was larger than .1 along with the existence of measurement error artifacts, meaning that the no-correction Hedges-Vevea's method could potentially perform quite stable even when statistical artifacts such as range restriction and measurement error exist. Although the present study did not specifically introduce the effect of range restriction on meta-analysis results into this investigation, it is possible that the estimation results could be still similar to each other for the Hedges-Vevea method and the two artifact correction methods.

Future Research

It is still undetermined whether or not artifact correction is effective and necessary in the field of meta-analysis. For example, Murphy (2003) mentioned that all the complexity in artifact correction models for estimation did not actually cause much difference among them, whereas the results of the bare-bones models turn out to be not very different from those that involve artifact correction. Aguinis et al. (2011) concurred that technical refinements in meta-analysis usually lead to very small substantive changes in the results and the subsequent conclusions. Lebreton, Scherer and James (2014) demonstrated their critical perspectives regarding the statistical values being used in artifact correction. They pointed out that the current pervasive culture of artifact correction encouraged a growing acceptance of studies that adopted artifact correction methods. However, some of the assumed reliability or range restriction values these studies incorporated were outdated and inappropriate, which led to an adjusted effect size estimate that was less creditable. In reality, true score is never known, so we can only use the comparison on real-world data to tell if the methods are very different from each other. If the differences are minimal, then does it really matter which analysis procedure we should use or to what extent a statistical assumption that a procedure requires is violated?

The current study revealed that neither the choice of artifact correction nor the choice of analysis procedure provided any significant improvement in the estimation results or in the research conclusions. It was the choice of the reliability estimates that provided noticeable differences in the analysis results. The impact of the violation of the assumption of independent reliability on the estimation results did not seem at all significant, which was predicted in Raju et al. (1998) study. The current findings to a certain degree support the general conclusions from other researchers such as Hall and Brannick (2002) as well as Raju et al. (1998). Admittedly, the core purpose of meta-analysis is to provide reliable, accurate, and stable estimates for the relation of variables, and thus any technical refinement that leads to an improvement in the estimation of correlations should be worthwhile (Köhler et al., 2015). But we should also seek for a logical balance between the practical gains and non-stop technical refinements for artifact correction. Perhaps it might be more worthwhile to devote additional efforts in improving the existing simple bare-bones procedures to be more resilient to statistical artifacts. This could be another research focus we can pursue to improve the use of meta-analysis in a wider field.

Some researchers (Lambert & Curlette, 1995; Oswald & Johnson, 1998) demonstrated that discrepancies between true ρ and its estimated ρ in a meta-analysis get larger if smaller within-study sample sizes and smaller numbers of effect sizes are included. Hunter and Schmidt (2004) dedicated a chapter to discussing the issue of second-order sampling error. If the number of sample studies is small, then even with appropriate application of the artifact correction formulas, there will still be non-trivial sampling error in the final meta-analysis results. Oswald and Johnson (1998) suggested that, compared with other inferential statistical procedures, artifact correction meta-analysis may actually be more negatively affected by assumption violations such as non-normal distributions in observed effect sizes. This may be because in meta-analysis there are many statistical corrections and each correction has its own set of assumptions (e.g., linearity, heteroscedasticity). All of these aforementioned researchers directed us to a potential remedy to improve the estimation results for artifact correction procedure, which is to select individual studies with a larger sample size and increase the number of studies included in a meta-analysis. The same idea can be applied to the bare-bones procedures to see if the differences between artifact correction procedures and no correction procedures will be narrowed when an infinite amount of studies with larger sample sizes are included in the meta-analysis. The impact of violation to normal distribution assumptions and pairwise independence assumption could also be tested under the condition of an increased amount of studies and increased sample sizes.

Limitations

This present research is inevitably subject to limitations. Although all efforts were made to reduce selection bias, it is quite possible that some relevant studies were not included in the current data. It is assumed that in the current study all the construct validities of job performance and organizational citizenship behaviors are hypothetically valid and consistent across sample studies using different measures. However, readers of this research should be aware that different measures might have represented different construct validities. Since construct validity is not the focus of the current study, caution should be paid due to biased population estimates potentially caused by these untenable assumptions.

Second, only limited types of reliability estimates were examined in the current study including coefficient alpha and intra-rater reliability. Other forms of reliability estimates, such as inter-rater, test-retest, and parallel forms, were not included due to their limited use in field practice especially in VG studies (Aguinis et al., 2011; Köhler et al., 2015). Hence, the conclusions from the current study should be interpreted with caution when these types of reliability estimates are the focus of a research. The interaction between reliability estimates and range restriction is also not the focus of the current study due to the fact that a large amount of studies have devoted efforts to this matter. However, Monte Carlo study results showed that the uncorrected alpha (internal consistency) suffered as a function of the selection ratio (range restriction) and the correlation between the test and the selection variable in both single and meta-analytic studies (Le & Schmidt, 2006; Li, 2013;).

In addition, the current research only tested selected artifact correction procedures, including the Hunter-Schmidt procedure and the RBNL procedure, on a limited amount of sample studies. The number of studies included in the current investigation can be considered as moderate. If the number of sample studies is small, the final meta-analysis results will suffer from second-sampling error (Hunter & Schmidt, 2004). Plus, Oswald and Johnson (1998) demonstrated that discrepancies between the population correlation ρ and its meta-analytic estimate ρ tend to get larger with smaller within-study sample sizes and with smaller numbers of effect sizes included in the meta-analysis. Therefore, the current research conclusion and implication should be interpreted and understood with caution. Data analyzed in the current study are limited to the Industrial and Organizational psychology field, hence generalizing the current conclusions to other research domains should be carefully considered. It is important to be cognizant of the fact that the findings here might not be replicated across other research fields.

Summary

After more than 20 years of methodological development, psychometric meta-analysis consumers are now surrounded by a myriad of artifact correction procedures and techniques. Most artifact correction methods are developed and applied on the condition of their respective assumptions, some of which may not always hold in research settings, thus cautions should be paid when a correction method is adopted.

The current study used the real-world data to compare the artifact correction procedures against the traditional non-correction meta-analysis procedures. The results revealed that neither the choice of artifact correction nor the choice of analysis procedure provided any significant difference in the estimation results or in the research conclusions, whereas it was the choice of the reliability estimates that provided noticeable differences in the analysis results. In addition, the impact of the violation to the assumption of independent reliability on the estimation results did not seem to be significant at all.

An adoption of any research and analysis method requires comprehensive understanding and careful application of this method in practice, and this ground rule is particularly critical for meta-analysis studies. As the statistical artifact correction procedures continue to develop, researchers should be aware that artifact correction should be applied appropriately, not mechanically. The ultimate goal of the methodology development for meta-analysis is to facilitate field research, rather than to devote excessive efforts in improving estimation techniques that yield no meaningful differences to meta-analysis consumers.

APPENDIX A

REFERENCES FOR CODED SAMPLE STUDIES

Aggarwal-Gupta, M., Vohra, N., & Bhatnagar, D. (2010). Perceived organizational support and organizational commitment: The meditational influence of psychological well-being. *Journal of Business and Management, 16,* 105-124.

Ahmad, Z. A., & Yekta, Z. A. (2010). Relationship between perceived organizational support, leadership behavior, and job satisfaction: An empirical study in Iran. *Intangible Capital, 6,* 162-184.

Al-Hussami, M. (2009). Predictors of nurses' commitment to health care organisations. *Australian Journal of Advanced Nursing, 26,* 36-48.

Al-Hussami, M., Saleh, M. Y. N., Abdalkader, R. H., & Mahadeen, A. I. (2011). Predictors of nursing faculty members' organizational commitment in governmental universities. *Journal of Nursing Management, 19,* 556-566.

Albrecht, S., & Travaglione, A. (2003). Trust in public-sector senior management. *International Journal of Human Resource Management, 14,* 76-92.

Alder, S. G., Noel, T. W., & Ambrose, M. L. (2006). Clarifying the effects of internet monitoring on job attitudes: The mediating role of employee trust. *Information & Management, 43,* 894-903.

Allen, D. G., Shore, L. M., & Griffeth, R. W. (2003). The role of perceived organizational support and supportive human resource practices in the turnover process. *Journal of Management, 29,* 99-118.

Allen, M. W. (1992). Communication and organizational commitment: Perceived organizational support as a mediating factor. *Communication Quarterly, 40,* 357-367.

Ambrose, M. L., & Schminke, M. (2003). Organization structure as a moderator of the relationship between procedural justice, interactional justice, perceived organizational support, and supervisory trust. *Journal of Applied Psychology, 88,* 295-305.

Anand, S., Vidyarthi, P. R., Liden, R. C., & Rousseau, D. M. (2010). Good citizens in poor-quality relationships: Idiosyncratic deals as a substitute for relationship quality. *Academy of Management Journal, 53,* 970-988.

Andrews, M. C., & Kacmar, K. M. (2001). Discriminating among organizational politics, justice, and support. *Journal of Organizational Behavior, 22,* 347-366.

Armstrong-Stassen, M. (1997). The effect of repeated downsizing and surplus designation on remaining managers: An exploratory study. *Anxiety, Stress & Coping, 10,* 377-384.

Armstrong-Stassen, M. (1998). Downsizing the federal government: A longitudinal study of managers' reactions. *Canadian Journal of Administrative Sciences, 15,* 310-321.

Armstrong-Stassen, M. (2003). Job transfer during organizational downsizing: A comparison of promotion and lateral transfers. *Group & Organization Management, 28,* 392-415.

Armstrong-Stassen, M. (2004). The influence of prior commitment on the reactions of layoff survivors to organizational downsizing. *Journal of Occupational Health Psychology, 9,* 46-60.

Armstrong-Stassen, M., & Schlosser, F. (2010). When hospitals provide HR practices tailored to older nurses, will older nurses stay? It may depend on their supervisor. *Human Resource Management Journal, 20,* 375-390.

Arshadi, N. (2011). The relationships of perceived organizational support (POS) with organizational commitment, in-role performance, and turnover intention: Mediating role of felt obligation. *Social and Behavioral Sciences, 30,* 1103-1108.

Asgari, A., Silong, A. D., Ahmad, A., & Samah, B. A. (2008). The relationship between transformational leadership behaviors, organizational justice, leader-member exchange, perceived organizational support, trust in management and organizational citizenship behaviors. *European Journal of Scientific Research, 23,* 227-242.

Aube, C., Rousseau, V., & Morin, E. M. (2007). Perceived organizational support and organizational commitment: The moderating effect of locus of control and work autonomy. *Journal of Managerial Psychology, 22,* 479-495.

Babakus, E., Cravens, D. W., Johnston, M., & Moncrief, W. C. (1996). Examining the role of organizational variables in the salesperson job satisfaction model. *Journal of Personal Selling & Sales Management, 16,* 33-46.

Bailey, L. C. (2007). *Refinement of an instrument to assess readiness for knowledge management* (Unpublished master's thesis). Air Force Institute of Technology, Wright-Patterson AFB, OH.

Bal, P. M., Chiaburu, D. S., & Jansen, P. G. W. (2010). Psychological contract breach and work performance: Is social exchange a buffer or an intensifier? *Journal of Managerial Psychology, 25,* 252-273.

Bambacas, M., & Bordia, P. (2009). Predicting different commitment components: The relative effects of how career development HRM practices are perceived. *Journal of Management & Organization, 15,* 224-240.

Baotham, S. (2011). Antecedent and consequence of job satisfaction and organizational commitment of Thai employees in RMUTT. *Review of Business Research, 11,* 89-98.

Baranik, L. E., Roling, E. A., & Eby, L. T. (2010). Why does mentoring work? The role of perceived organizational support. *Journal of Vocational Behavior, 76,* 366-373.

Barnett, C. M. (1996). *The effects of creativity and perceived support on organizational commitment* (Unpublished doctoral dissertation). California State University, Long Beach.

Behson, S. J. (2002). Which dominates? The relative importance of work-family organizational support and general organizational context on employee outcomes. *Journal of Vocational Behavior, 61,* 53-72.

Bernal, B. W. (1997). *Repatriation adjustment: A study of American employees after return from overseas assignments* (Unpublished doctoral dissertation). California State University, Long Beach.

Bettencourt, L. A., Gwinner, K. P., & Meuter, M. L. (2001). A comparison of attitude, personality, and knowledge predictors of service-oriented organizational citizenship behaviors. *Journal of Applied Psychology, 86,* 29-41.

Bhanthumnavin, D. (2003). Perceived social support from supervisor and group members' psychological and situational characteristics as predictors of subordinate performance in Thai work units. *Human Resource Development Quarterly, 14,* 79-97.

Bishop, J. W., Goldsby, M. G., & Neck, C. P. (2002). Who goes? Who cares? Who stays? Who wants to? The role of contingent workers and corporate layoff practices. *Journal of Managerial Psychology, 17,* 298-315.

Bishop, J. W., Scott, K. D., & Burroughs, S. M. (2000). Support, commitment, and employee outcomes in a team environment. *Journal of Management, 26,* 1113-1132.

Bishop, J. W., Scott, K. D., Goldsby, M. G., & Cropanzano, R. (2005). A construct validity study of commitment and perceived support variables: A multifoci approach across different team environments. *Group & Organization Management, 30,* 153-180.

Bissoonauth, B. (2003). *Virtual project work: Investigating critical success factors of virtual project performance* (Unpublished doctoral dissertation). Concordia University, Irvine, CA.

Blakely, G. L., Andrews, M. C., & Fuller, J. (2003). Are chameleons good citizens? A longitudinal study of the relationship between self-monitoring and organizational citizenship behavior. *Journal of Business and Psychology, 18,* 131-144.

Blau, G. (2001). Testing the discriminant validity of occupational entrenchment. *Journal of Occupational and Organizational Psychology, 74,* 85-93.

Blau, G., Tatum, D. S., & Ward-Cook, K. (2003b). Correlates of work exhaustion for medical technologists. *Journal of Allied Health, 32,* 148-157.

Bogle, C. A. (2010). *An integrative model of the feedback environment* (Unpublished doctoral dissertation). Florida Institute of Technology, Melbourne.

Bradley, B. L. (1997). *The relationships between perceived organizational support, institutional orientation and affective, continuance, and normative commitment in the Canadian military* (Unpublished doctoral dissertation). University of Calgary, Calgary, Canada.

Bradley, J. R., & Cartwright, S. (2002). Social support, job stress, health, and job satisfaction among nurses in the United Kingdom. *International Journal of Stress Management, 9,* 163-182.

Brandes, P., Dharwadkar, R., & Wheatley, K. (2004). Social exchanges within organizations and work outcomes: The importance of local and global relationships. *Group & Organization Management, 29,* 276-301.

Brown, D. J., Ferris, D. L., Heller, D., & Keeping, L. M. (2007). Antecedents and consequences of the frequency of upward and downward social comparisons at work. *Organizational Behavior and Human Decision Processes, 102,* 59-75.

Brown, L. (1992). *Stress and burnout in a sample of Canadian federal public servants* (Unpublished doctoral dissertation). Carleton University, Ottawa, Canada.

Burnett, M. P. (1997). *Perceived organizational support, organizational dependability, and organizational commitment* (Unpublished doctoral dissertation). California State University, Long Beach.

Burnstad, L. D. (1994). *Employer offered family assistance provisions: The moderating effect of perceived organizational support on organizational commitment* (Unpublished doctoral dissertation). University of Alberta, Edmonton, Canada.

Butts, M. M., Vandenberg, R. J., DeJoy, D. A., Schaffer, B. S., & Wilson, M. G. (2009). Individual's reactions to high-involvement work processes: Investigating the role of empowerment and perceived organizational support. *Journal of Occupational Health Psychology, 14,* 122-136.

Byrne, Z. S., & Hochwarter, W. A. (2006). I get by with a little help from my friends: The interaction of chronic pain and organizational support on performance. *Journal of Occupational Health Psychology, 11,* 215-227.

Byrne, Z. S., & Hochwarter, W. A. (2008). Perceived organizational support and performance: Relationships across levels of organizational cynicism. *Journal of Managerial Psychology, 23,* 54-72.

Byrne, Z., Pitts, V., Chiaburu, D., & Steiner, Z. (2011). Managerial trustworthiness and social exchange with the organization. *Journal of Managerial Psychology, 26,* 108-122.

Cable, D. M., & DeRue, D. S. (2002). The convergent and discriminant validity of subjective fit perceptions. *Journal of Applied Psychology, 87,* 875-884.

Camerman, J., Cropanzano, R., & Vandenberghe, C. (2007). The benefits of justice for temporary workers. *Group & Organization Management, 32,* 176-207.

Campbell, D. J., & Campbell, K. M. (2003). Global versus facet predictors of intention to quit: Differences in a sample of male and female Singaporean managers and non-managers. *International Journal of Human Resource Management, 14,* 1152-1177.

Canipe, J. S. (2006). *Relationships among trust, organizational commitment, perceived organizational support, and turnover intentions* (Unpublished doctoral dissertation). Alliant International University, San Diego, CA.

Capon, J., Chernyshenko, O. S., & Stark, S. (2007). Applicability of civilian retention theory in the New Zealand military. *New Zealand Journal of Psychology, 36,* 50-56.

Carlson, S. C. (2005). *Ethical leadership: Influences of ethical climate, perceived organizational support, and perceived leader integrity* (Unpublished doctoral dissertation). Nova Southeastern University, Fort Lauderdale, FL.

Carver, L. L. (2008). *Organizational commitment and generational differences in nursing faculty* (Unpublished doctoral dissertation). University of Nevada, Las Vegas.

Carver, L., Candela, L., & Guiterrez, A. P. (2011). Survey of generational aspects of nurse faculty organizational commitment. *Nursing Outlook, 59,* 137-148.

Casper, W. J., Martin, J. A., Buffardi, L. C., & Erdwins, C. J. (2002). Work-family conflict, perceived organizational support, and organizational commitment among employed mothers. *Journal of Occupational Health Psychology, 7,* 99-108.

Casper, W. J., Weltman, D., & Kwesiga, E. (2007). Beyond family-friendly: The construct and measurement of singles-friendly culture. *Journal of Vocational Behavior, 70,* 478-501.

Chambel, M. J., & Sobral, F. (2011). Training is an investment with return in temporary workers: A social exchange perspective. *Career Development International, 16,* 161-177.

Chan, D. (2001). Method effects of positive affectivity, negative affectivity, and impression management in self-reports of work attitudes. *Human Performance, 14,* 77-96.

Chan, S., & Jepsen, D. (2009). Organisational justice: A hospitality shift worker contextual perspective. In *2009 International Employment Relations Association (IERA) Conference—Book of Proceedings* (pp. 1-18). Bangkok, Thailand: International Employment Relations Association.

Chau, S. L. (2008). *Examining the emotional labor process: A moderated model of emotional labor and its effects on job performance and turnover* (Unpublished doctoral dissertation). University of Akron, Akron, OH.

Chen, Z. X., Aryee, S., & Lee, C. (2005). Test of a mediation model of perceived organizational support. *Journal of Vocational Behavior, 66,* 457-470.

Cheung, F. Y. (2005). *Employee reciprocation to psychological contract fulfillment: A mediating role focusing on multiple exchange perspectives (China)* (Unpublished doctoral dissertation). Hong Kong Polytechnic University.

Chiang, H., Han, T., & Chuang, J. (2011). The relationship between high-commitment HRM and knowledgesharing behavior and its mediators. *International Journal of Manpower, 32,* 604-622.

Cho, J., & Treadway, D. C. (2011). Organizational identification and perceived organizational support as mediators of the procedural justice-citizenship behaviour relationship: A cross-cultural constructive replication. *European Journal of Work and Organizational Psychology, 20,* 631-653.

Cho, S., Johanson, M. M., & Guchait, P. (2009). Employee's intent to leave: A comparison of determinants of intent to leave versus intent to stay. *International Journal of Hospitality Management, 28,* 374-381.

Choi, J. N. (2006). Multilevel and cross-level effects of workplace attitudes and group member relations on interpersonal helping behavior. *Human Performance, 19,* 383-402.

Christopher, J. (2000). *Organizational justice and employee evaluations of benefit programs* (Unpublished doctoral dissertation). Wayne State University, Detroit, MI.

Clark, S. W. (2003). *The development of an integrated measure of readiness for change instrument and its application on ASC/PK* (Unpublished master's thesis). Air Force Institute of Technology, Wright-Patterson AFB, OH.

Cleveland, J. N., & Shore, L. M. (1992). Self-perspectives and supervisory perspectives on age and work attitudes and performance. *Journal of Applied Psychology, 77,* 469-484.

Colakoglu, U., Culha, O., & Atay, H. (2010). The effects of perceived organisational support on employees' affective outcomes: Evidence from the hotel industry. *Tourism and Hospitality Management, 16,* 125-126.

Connell, J., Ferres, N., & Travaglione, T. (2003). Engendering trust in manager-subordinate relationships: Predictors and outcomes. *Personnel Review, 32,* 569-587.

Coyle-Shapiro, J., & Kessler, I. (2000). Consequences of the psychological contract for the employment relationship: A large scale survey. *Journal of Management Studies, 37,* 903-930.

Cropanzano, R., Howes, J. C., Grandey, A. A., & Toth, P. (1997). The relationship of organizational politics and support to work behaviors, attitudes, and stress. *Journal of Organizational Behavior, 22,* 159-180.

Daly, C. L. (1999). *Explaining the extra-role behavior of part-time and full-time workers: A matter of support?* (Unpublished doctoral dissertation). Michigan State University, East Lansing.

Darolia, C. R., Kumari, P., & Darolia, S. (2010). Perceived organizational support, work motivation, and organizational commitment as determinants of job performance. *Journal of the Indian Academy of Applied Psychology, 36,* 69-78.

Dawley, D. D., Andrews, M. C., & Bucklew, N. S. (2008). Mentoring, supervisor support, and perceived organizational support: What matters most? *Leadership & Organization Development Journal, 29,* 235-247.

Dawson, C. L. (1996). *Dispositional and attitudinal explanations of counter productivity in the workplace* (Unpublished doctoral dissertation). University of California, Berkeley.

DeConinck, J. B. (2010). The effect of organizational justice, perceived organizational support, and perceived supervisor support on marketing employees' level of trust. *Journal of Business Research, 63,* 1349-1355.

DeConinck, J. B., & Johnson, J. T. (2009). The effects of perceived supervisor support, perceived organizational support, and organizational justice on turnover among salespeople. *Journal of Personal Selling & Sales Management, 29,* 333-350.

Desrosiers, E. I. (2003). *Telework and work attitudes: The relationship between telecommuting and employee job satisfaction, organizational commitment, perceived organizational support, and perceived co-worker support* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.

Dion, M. J. (2006). *The impact of workplace incivility and occupational stress on job satisfaction and turnover intention of acute care nurses* (Unpublished doctoral dissertation). University of Connecticut, Storrs.

Dobreva-Martinova, T. (2002). *Occupational stress in the Canadian forces: Its association with individual and organizational well-being* (Unpublished doctoral dissertation). Carleton University, Ottawa, Canada.

Dourigan, J. P. (1995). *The effects of organizational change on professionals' and organizationals' resistance to work change, organizational loss, and commitment levels* (Unpublished doctoral dissertation). California State University, Long Beach.

Draves, P. R. (2003). *An examination of potential moderating effects of personality on the relationship between job attitudes and organizational citizenship behaviors* (Unpublished doctoral dissertation). University of South Florida, Tampa.

Duffy, R. D., & Lent, R. W. (2009). Test of a social cognitive model of work satisfaction in teachers. *Journal of Vocational Behavior, 75,* 212-223.

Duke, A. B., Goodman, J. M., Treadway, D. C., & Breland, J. W. (2009). Perceived organizational support as a moderator of emotional labor/outcomes relationships. Journal of *Applied Social Psychology, 39,* 1013-1034.

Dulac, T., Coyle-Shapiro, J. A. M., Henderson, D. J., & Wayne, S. J. (2008). Not all responses to breach are the same: The interconnection of social exchange and psychological contract processes in organizations. *Academy of Management Journal, 51,* 1079-1098.

Dupre, K. E., & Day, A. L. (2007). The effects of supportive management and job quality on the turnover intentions and health of military personnel. *Human Resource Management, 46,* 185-201.

Edwards, M. R. (2009). HR, perceived organisational support, and organisational identification: An analysis after organizational formation. *Human Resource Management Journal, 19,* 91-115.

Eisenberger, R., Armeli, S., Rexwinkel, B., Lynch, P. D., & Rhoades, L. (2001). Reciprocation of perceived organizational support. *Journal of Applied Psychology, 86,* 42-51.

Eisenberger, R., Cummings, J., Armeli, S., & Lynch, P. (1997). Perceived organizational support, discretionary treatment, and job satisfaction. *Journal of Applied Psychology, 82,* 812-820.

Eisenberger, R., Fasolo, P., & Davis-LaMastro, V. (1990). Perceived organizational support and employee diligence, commitment, and innovation. *Journal of Applied Psychology, 75,* 51-59.

Eisenberger, R., Rhoades, L., & Cameron, J. (1999). Does pay for performance increase or decrease perceived self-determination and intrinsic motivation? *Journal of Personality and Social Psychology, 77,* 1026-1040.

El Akremi, A., Vandenberghe, C., & Camerman, J. (2010). The role of justice and social exchange relationships in workplace deviance: Test of a mediated model. *Human Relations, 63,* 1687-1717.

Ellis, B. A. (2005). *Employees' reaction to the introduction of enterprise resource planning system* (Unpublished doctoral dissertation). Regent University, Virginia Beach, VA.

Erdogan, B., Kraimer, M. L., & Liden, R. C. (2004). Work value congruence and intrinsic career success: The compensatory roles of leader-member exchange and perceived organizational support. *Personnel Psychology, 57,* 305-332.

Erdwins, C. J., Buffardi, L. C., Casper, W. J., & O'Brien, A. S. (2001). The relationship of women's role strain to social support, role satisfaction, and self-efficacy. *Family Relations, 50,* 230-238.

Erickson, R. A. (2007). *Here today but what about tomorrow? Reducing the attrition of downsizing survivors by increasing their organizational commitment* (Unpublished doctoral dissertation). Northwestern University, Evanston, IL.

Farh, J. L., Hackett, R. D., & Liang, J. (2007). Individual-level cultural values as moderators of perceived organizational support–employee outcome relationships in China: Comparing the effects of power distance and traditionality. *Academy of Management Journal, 50,* 715-729.

Farrell, M. A., & Oczkowski, E. (2009). Service worker customer orientation, organisation/job fit and perceived organisational support. *Journal of Strategic Marketing, 17,* 149-167.

Fausz, A. T. (1995). *Factors related to supervisory ratings of employees' customer service orientation* (Unpublished doctoral dissertation). University of Tennessee, Knoxville.

Feng, W. C., & Angeline, T. (2010). Turnover intention and job hopping behaviour of music teachers in Malaysia. *African Journal of Business Management, 4,* 425-434.

Ferres, N., Connell, J., & Travaglione, A. (2004). Co-worker trust as a social catalyst for constructive employee attitudes. *Journal of Managerial Psychology, 19,* 608-622.

Ferres, N., Connell, J., & Travaglione, A. (2005). The effect of future redeployment on organizational trust. *Strategic Change, 14,* 77-91.

Filipova, A. A. (2008). *Perceived organizational support and ethical work climates as predictors of turnover intention of licensed nurses in skilled nursing facilities* (Unpublished doctoral dissertation). Western Michigan University, Kalamazoo.

Florey, A. T. (1998). *Decision to make an accommodation request: Theory and evidence from the perspective of employees with disabilities* (Unpublished doctoral dissertation). University of Texas Arlington.

Fodchuk, K. M. (2009). *Organizational justice perceptions in China: Development of the Chinese organizational justice scale* (Unpublished doctoral dissertation). Old Dominion University, Norfolk, VA.

Foley, S., Hang-yue, N., & Loi, R. (2006). How do cultural types affect work-related attitudes? The mediating role of perceived organisational support. *International Journal of Employment Studies, 14,* 37-62.

Fortunato, V. J. (2004). A comparison of the construct validity of three measures of negative affectivity. *Educational and Psychological Measurement, 64,* 271-289.

Fu, F., Bolander, W., & Jones, E. (2009). Managing the drivers of organizational commitment and salesperson effort: An application of Allen's three-component model. *Journal of Marketing Theory and Practice, 17,* 335-350.

Fuller, J. B., Hester, K., Barnett, T., Frey, L., & Relyea, C. (2006). Perceived organizational support and perceived external prestige: Predicting organizational attachment for university faculty, staff, and administrators. *Journal of Social Psychology, 146,* 327-347.

Gakovic, A., & Tetrick, L. E. (2003). Perceived organizational support and work status: A comparison of the employment relationships of part-time and full-time employees attending university classes. *Journal of Organizational Behavior, 24,* 649-666.

Gandolfo, C. J. (2006). *The role of leader member exchange theory and multi-rater feedback on evaluating the performance appraisal process* (Unpublished doctoral dissertation). Southern Illinois University at Carbondale.

Garonzik, R. (1993). *A multivariate study of expatriate managerial performance: The effects of organizational, environmental, and personal factors* (Unpublished doctoral dissertation). Boston University, Boston, MA.

Gautam, T., van Dick, R., & Wagner, U. (2001). Organizational commitment in Nepalese settings. *Asian Journal of Social Psychology, 4,* 239-248.

Gavino, M. (2005). *Understanding the impact of HR practices on employee attitudinal and behavioral outcomes: The role of POS* (Unpublished doctoral dissertation). University of Illinois at Chicago.

Gibney, R., Zagenczyk, T. J., & Masters, M. F. (2009). The negative aspects of social exchange: An introduction to perceived organizational obstruction. *Group & Organization Management, 34,* 665-697.

Gibson, J. L. (2006). *Employees and work-life resources: Influences on attraction, spillover, and commitment* (Unpublished doctoral dissertation). George Mason University, Fairfax, VA.

Goeddeke, F. X. (2005). *Organizational trust as a mediator of individual union voting intentions in faculty union certification elections* (Unpublished doctoral dissertation). University of Florida, Gainesville.

Goeddeke, F. X., Jr., & Kammeyer-Mueller, J. D. (2010). Perceived support in a dual organizational environment: Union participation in a university setting. *Journal of Organizational Behavior, 31,* 65-83.

Gosserand, R. H. (2003). *An examination of individual and organizational factors relating to emotional labor* (Unpublished doctoral dissertation). Louisiana State University, Baton Rouge.

Griffin, M. L. (2001). Job satisfaction among detention officers: Assessing the relative contribution of organizational climate variables. *Journal of Criminal Justice, 29,* 219-232.

Griffin, M. L., & Hepburn, J. R. (2005). Side-bets and reciprocity as determinants of organizational commitment among correctional officers. *Journal of Criminal Justice, 33,* 611-625.

Gyekye, S. A., & Salminen, S. (2009). Perceived organizational support: An African perspective. *Journal of Applied Social Psychology, 39,* 2651-2668.

Haar, J. M., & Spell, C. S. (2004). Programme knowledge and value of work-family practices and organizational commitment. *International Journal of Human Resource Management, 15,* 1040-1055.

Hamwi, G. A. (2010). *The mediating role of ethical decision making in the relationship between job characteristics and job outcomes: An examination of business-to-business salespeople* (Unpublished doctoral dissertation). Georgia State University, Atlanta.

Hannigan, M. A. (2003). *Family supportive organization perceptions and the relationship between family-related benefits and organizational attitudes* (Unpublished doctoral dissertation). Wayne State University, Detroit, MI.

Harlos, K. P., & Axelrod, L. J. (2005). Investigating hospital administrators' experiences of workplace mistreatment. *Canadian Journal of Behavioural Science, 37,* 262-272.

Harris, K. A. (1995). *Variables related to supervisory confrontation and referral of employees to employee assistance programs* (Unpublished doctoral dissertation). Wayne State University, Detroit, MI.

Harris, R. B., Harris, K. J., & Harvey, P. (2007). A test of competing models of the relationships among perceptions of organizational politics, perceived organizational support, and individual outcomes. *Journal of Social Psychology, 147,* 631-655.

Hawkins, A. S. (2005). *The relationship of perceived learning culture to perceived organizational support and affective commitment* (Unpublished doctoral dissertation). Alliant International University, San Diego, CA.

Hawkins, W. D. (1998). *Predictors of affective organizational commitment among high school principals* (Unpublished doctoral dissertation). Virginia Polytechnic Institute and State University, Blacksburg.

Herda, D. N. (2011). *Social exchange and post-employment citizenship: Evidence from public accounting* (Unpublished doctoral dissertation). University of Texas at Arlington.

Hochwarter, W. A., Witt, L. A., Treadway, D. C., & Ferris, G. R. (2006). The interaction of social skill and organizational support on job performance. *Journal of Applied Psychology, 91,* 482-489.

Howes, J. C., Cropanzano, R., Grandey, A. A., & Mohler, C. J. (2000).Who is supporting whom? Quality team effectiveness and perceived organizational support. *Journal of Quality Management, 5,* 207-223.

Hui, C., Lee, C., & Rousseau, D. M. (2004). Employment relationships in China: Do workers relate to the organization or to people? *Organization Science, 15,* 232-240.

Hui, C., Wong, A., & Tjosvold, D. (2007). Turnover intention and performance in China: The role of positive affectivity, Chinese values, perceived organizational support and constructive controversy. *Journal of Occupational and Organizational Psychology, 80,* 735-751.

Hung, H., & Wong, Y. H. (2007). The relationship between employer endorsement of continuing education and training and work and study performance: A Hong Kong case study. *International Journal of Training and Development, 11,* 295-313.

Hutchison, S. (1997). A path model of perceived organizational support. *Journal of Social Behavior and Personality, 12,* 159-174.

Hutchison, S., & Garstka, M. L. (1996). Sources of perceived organizational support: Goal setting and feedback. *Journal of Applied Social Psychology, 26,* 1351-1366.

Hutchison, S., Valentino, K. E., & Kirkner, S. L. (1998). What works for the gander does not work as well for the goose: The effects of leader behavior. *Journal of Applied Social Psychology, 28,* 171-182.

Hyland, M. M. (1999). *Flexibility in work arrangements: How availability, preferences and use affect business outcomes* (Unpublished doctoral dissertation). Rutgers, State University of New Jersey–New Brunswick.

James, M. (2005). *Antecedents and consequences of cynicism in organizations: An examination of the potential positive and negative effects on school systems* (Unpublished doctoral dissertation). Florida State University, Tallahassee.

Jawahar, I. M., & Carr, D. (2007). Conscientiousness and contextual performance: The compensatory effects of perceived organizational support and leader-member exchange. *Journal of Managerial Psychology, 22,* 330-349.

Jefferson, D. J. (1990). *Commitment to work and the organization: The role of social exchange theory* (Unpublished doctoral dissertation). University of Delaware, Newark.

Jing-zhou, P., & Wen-xia, Z. (2011). Under dual perspective of work exchange and social exchange. *Nankai Business Review International, 2,* 402-417.

Jing-zhou, P., Xiao-xue, Z., & Xia-qing, Z. (2010). The role of leadership between the employees and the organization: A bridge or a ravine? An empirical study from China. *Journal of Management & Marketing Research, 5,* 51-14.

Johanson, M. M., & Cho, S. (2009). Uncovering the link between organizational behaviors and employment status in the U.S. hotel industry. *Journal of Human Resources in Hospitality & Tourism, 8,* 184-198.

Jones, B., Flynn, D. M., & Kelloway, E. K. (1995). Perception of support from the organization in relation to work stress, satisfaction, and commitment. In S. L. Saunter & L. R. Murphy (Eds.), *Organizational risk factors for job stress* (pp. 41-52). Washington, DC: American Psychological Association.

Junak, J. A. (2007). *Effects of perceived organizational support on turnover intention through job satisfaction and organizational commitment* (Unpublished master's thesis). Roosevelt University, Chicago, IL.

Jung, H. H. (2003). *An international study of organizational change: A simultaneous analysis of process, context, and individual attributes* (Unpublished master's thesis). Air Force Institute of Technology, Wright-Patterson AFB, OH.

Kang, B., Twigg, N. W., & Hertzman, J. (2009). An examination of social support and social identity factors and their relationship to certified chefs' burnout. International Journal of *Hospitality Management, 29,* 168-176.

Keeton, K. E. (2008). *An extension of the UTAUT model: How organizational factors and individual differences influence technology acceptance* (Unpublished doctoral dissertation). University of Houston, Houston, TX.

Kehrli, S. E. (2008). *Repatriates' organizational commitment and turnover intentions in relation to availability of HR practices and perceptions of organizational and supervisory support* (Unpublished doctoral dissertation). Alliant International University, San Diego, CA.

Keough, S. M. (2006). *A social exchange perspective of intention to quit* (Unpublished doctoral dissertation). Mississippi State University, Starkville.

Kernan, M. C., Watson, S., Chen, F. F., & Kim, T. G. (2011). How cultural values affect the impact of abusive supervision on worker attitudes. *Cross Cultural Management, 18,* 464-484.

Kerr, V. O. (2005). *Influence of perceived organizational support, organizational commitment, and professional commitment on turnover intentions of healthcare professionals in Jamaica* (Unpublished doctoral dissertation). Nova Southeastern University, Fort Lauderdale.

Khazanchi, S. (2005). *A ''social exchange'' model of creativity* (Unpublished doctoral dissertation). University of Cincinnati, Cincinnati, OH.

Khurram, S. (2009). Perceived organizational support, antecedents and consequences: Proposing and testing a model in a public sector university of Pakistan. *South Asian Journal of Management, 16,* 7-26.

Kim, W. G., Leong, J. K., & Lee, Y. K. (2005). Effect of service orientation on job satisfaction, organizational commitment, and intention of leaving in a casual dining chain restaurant. *International Journal of Hospitality Management, 24,* 171-193.

Kirsh, M. C. (2011). *Job satisfaction and intention to quit: A study of health care employees* (Unpublished master's thesis). Northern Illinois University, DeKalb.

Kittredge, A. A. (2010). *Predicting work and organizational engagement with work and personal factors* (Unpublished master's thesis). San Jose State University, San Jose, CA.

Klammer, J. D. (1997). *Speaking up in the workplace: Modeling the implications of procedural justice, perceived organizational support, being heard, and tokenism in the relationship between voice and civic virtue* (Unpublished doctoral dissertation). University of Calgary, Calgary, Canada.

Koenig, L. D. (1992). *Perceived organizational support and employee commitment, work effort, and turnover intent* (Unpublished doctoral dissertation). California State University, Long Beach.

Kogan, L. (2004). *Mediating effects of affective commitment and perceived organizational support on the relationship between procedural justice and organizational citizenship behavior* (Unpublished doctoral dissertation). Alliant International University, San Diego, CA.

Konik, J. A. (2005). *Harassment as a system for policing traditional gender norms in the workplace: The structure and process of sexual harassment and heterosexist harassment* (Unpublished doctoral dissertation). University of Michigan, Ann Arbor.

Kraimer, M. L. (1999). *Expatriate adjustment: Integrating the stress management and sensemaking perspectives* (Unpublished doctoral dissertation). University of Illinois at Chicago.

Kuvaas, B. (2008). An exploration of how the employee-organization relationship affects the linkage between perception of developmental human resource practices and employee outcomes. *Journal of Management Studies, 45,* 1-25.

Kyei-Poku, I. A. (2009). *Toward a better understanding of fairness in the workplace: The role of psychological uncertainty and burnout* (Unpublished doctoral dissertation). University of Western Ontario, London, Canada.

Ladd, D., & Henry, R. A. (2000). Helping coworkers and helping the organization: The role of support perceptions, exchange ideology, and conscientiousness. *Journal of Applied Social Psychology, 30,* 2028-2049.

Ladebo, O. (2004). Employees' personal motives for engaging in citizenship behavior: The case of workers in Nigeria's agriculture industry. *Current Research in Social Psychology, 9* (no pagination specified).

Ladebo, O. J. (2009). Emotional exhaustion and strain reactions: Perceived organisational support as a moderator. *South African Journal of Psychology, 39,* 46-58.

Ladebo, O. J., Abubakar, B. Z., & Adamu, C. O. (2011). Nigerian agriculture workers' outcomes from perceived organisational support and protestant work ethics: Job satisfaction as a mediator. *South African Journal of Industrial Psychology, 37,* 89-96.

LaMastro, V. A. (1995). *Factors affecting professional commitment, organizational commitment, and perceived organizational support in a sample of educators* (Unpublished doctoral dissertation). University of Delaware, Newark.

Lapalme, M.-E`. Tremblay, M., & Simard, G. (2009). The relationship between career plateauing, employee commitment and psychological distress: The role of organizational and supervisor support. *The International Journal of Human Resource Management, 20,* 1132-1145.

Laschinger, H. K., Purdy, N., Cho, J., & Almost, J. (2006). Antecedents and consequences of nurse managers' perceptions of organizational support. *Nursing Economics, 24,* 20-29.

Lavelle, J. J., McMahan, G. C., & Harris, C. M. (2009). Fairness in human resource management, social exchange relationships, and citizenship behavior: Testing linkages of the target similarity model among nurses in the United States. *The International Journal of Human Resource Management, 20,* 2419-2434.

Lazarova, M., & Caligiuri, P. (2001). Retaining repatriates: The role of organizational support practices. *Journal of World Business, 36,* 389-401.

Lee, J., & Peccei, R. (2007). Perceived organizational support and affective commitment: The mediating role of organization-based self-esteem in the context of job insecurity. *Journal of Organizational Behavior, 28,* 661-685.

Lee, J., & Peccei, R. (2011). Discriminant validity and interaction between perceived organizational support and perceptions of organizational politics: A temporal analysis. *Journal of Occupational and Organizational Psychology, 84,* 686-702.

Lee, S. (2004). *Organizational commitment and job satisfaction: The effect of alternative officing strategies on teleworkers' organizational behavior* (Unpublished doctoral dissertation). Michigan State University, East Lansing.

Lent, R. W., Nota, L., Soresi, S., Ginevra, M. C., Duffy, R. D., & Brown, S. D. (2011). Predicting the job and life satisfaction of Italian teachers: Test of a social cognitive model. *Journal of Vocational Behavior, 79,* 91-97.

Li, P. (1999). *Effects of sales force control systems on salesperson job outcomes: A psychological climate and contingency perspective* (Unpublished doctoral dissertation). University of Missouri–Columbia.

Liao, H., Joshi, A., & Chuang, A. (2004). Sticking out like a sore thumb: Employee dissimilarity and deviance at work. *Personnel Psychology, 57,* 969-1000.

Liao, H., Toya, K., Lepak, D. P., & Hong, Y. (2009). Do they see eye to eye? Management and employee perspectives of high-performance work systems and influences processes on service quality. *Journal of Applied Psychology, 94,* 371-391.

Liao, P. (2011). Linking work-family conflict to job attitudes: The mediating role of social exchange relationships. *The International Journal of Human Resource Management, 22,* 2965-2980.

Liden, R. C., Wayne, S. J., Kraimer, M. L., & Sparrowe, R. (2003). The dual commitments of contingent workers: An examination of contingents' commitments to the agency and the organization. *Journal of Organizational Behavior, 24,* 609-625.

Lilly, J. D., & Virick, M. (2006). The effect of personality on perceptions of justice. *Journal of Managerial Psychology, 21,* 438-458.

Lin, J. S., & Lin, S. (2011). Moderating effect of organizational climate on the relationship of organizational support and service-oriented organizational citizenship behaviors. *African Journal of Business Management, 5,* 582-595.

Liu, W. (2004). *Perceived organizational support: Linking human resources management practices with important work outcomes* (Unpublished doctoral dissertation). University of Maryland, College Park.

Liu, Y. W. (2009). Perceived organizational support and expatriate organizational citizenship behavior: The mediating role of affective commitment towards the parent company. *Personnel Review, 38,* 307-19.

Loi, C. H. (2006). *The roles of perceived organizational support, social exchange, and economic exchange in employment relationships in China* (Unpublished doctoral dissertation). Chinese University of Hong Kong.

Loi, R., & Ngo, H.-Y. (2010). Mobility norms, risk aversion, and career satisfaction of Chinese employees. *Asia Pacific Journal of Management, 27,* 237-255.

Loi, R., Hang-Yue, N., & Foley, S. (2006). Linking employees' justice perceptions to organizational commitment and intention to leave: The mediating role of perceived organizational support. *Journal of Occupational and Organizational Psychology, 79,* 101-120.

Lopez, Y. P. (2008). *Re-defining psychological contracts: A network of relationships* (Unpublished doctoral dissertation). Texas A&M University, College Station.

Lynch, P., Eisenberger, R., & Armeli, S. (1999). Perceived organizational support: Inferior-versus-superior performance by wary employees. *Journal of Applied Psychology, 84,* 467-483.

Maertz, C. P., Griffeth, R. W., Campbell, N. S., & Allen, D. G. (2007). The effects of perceived organizational support and perceived supervisor support on employee turnover. *Journal of Organizational Behavior, 28,* 1059-1075.

Malatesta, R. M. (1996). *Understanding the dynamics of organizational and supervisory commitment using a social exchange framework* (Unpublished doctoral dissertation). Wayne State University, Detroit, MI.

Mallette, C. (2005). *Theoretical and nursing implications of employment patterns and social exchange theory: Impact on the individual, organization, and the profession* (Unpublished doctoral dissertation). University of Toronto, Toronto, Canada.

Marquardt, P. A. (2002). *Organizational commitment in long-term outsourcing arrangements: An investigation of a new breed of dual allegiance* (Unpublished doctoral dissertation). Kansas State University, Manhattan.

Martinez, P. G. (2002). *Paternalism in contemporary organizations: A theory and test of organizational benevolence and decision-making control* (Unpublished doctoral dissertation). University of California, Irvine.

Masterson, S. S. (1998). *A ''trickle-down'' model of organizational justice: Relating employee and customer perceptions of and reactions to organizational fairness* (Unpublished doctoral dissertation). University of Maryland, College Park.

Masterson, S. S., Lewis, K., Goldman, B. M., & Taylor, M. S. (2000). Integrating justice and social exchange: The differing effects of fair procedures and treatment on work relationships. *Academy of Management Journal, 43,* 738-748.

Matsuda, Y., Pierce, J. L., & Ishikawa, R. (2011). Development and validation of the Japanese version of the organization-based self-esteem scale. *Journal of Occupational Health, 53,* 188-196.

Mauseth, K. B. (2008). *The influence of perceived organizational support and school culture on positive workplace outcomes for teachers in private schools* (Unpublished doctoral dissertation). Seattle Pacific University, Seattle, WA.

McCook, K. D. (2002). *Organizational perceptions and their relationship to job attitudes, effort, performance, and organizational citizenship behavior* (Unpublished doctoral dissertation). Louisiana State University, Baton Rouge.

McIntyre, K. S. (2005). *Understanding a climate for customer service* (Unpublished doctoral dissertation). Colorado State University, Fort Collins.

Meador, J. L. (2001). *Job satisfaction, perceived organizational support, and organizational commitment: Implications for teacher turnover in small rural schools* (Unpublished doctoral dissertation). Stephen F. Austin University, Nacogdoches, TX.

Meikle, S. (2011). *The physical work environment in relation to the physical and psychological symptoms of sick building syndrome* (Unpublished master's thesis). University of Waikato, Hamilton, New Zealand.

Menassa, J. (2007). *Valuing diversity: An examination of diversity climate, work attitudes, and employee racioethnicities* (Unpublished master's thesis). Concordia University, Irvine, CA.

Messer, C. O. (2006). *Resistance to change in the community college: The influence of participation, open communication, perceived organizational support, and organizational commitment* (Unpublished doctoral dissertation). University of Oklahoma, Norman.

Miao, R., & Kim, H. (2010). Perceived organizational support, job satisfaction and employee performance: A Chinese empirical study. *Journal of Service Science & Management, 3,* 257-264.

Moffitt, K. R. (2009). *Modeling the relationship between traumatic caregiver stress and job satisfaction, organizational commitment, and turnover intentions* (Unpublished doctoral dissertation). University of Memphis, Memphis, TN.

Moideenkutty, U. (2000). *Correlates and outcomes of organizational citizenship behavior directed toward the organization, the supervisor, and co-workers: A social exchange perspective* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.

Moideenkutty, U., Blau, G., Kumar, R., & Nalakath, A. (2001). Perceived organisational support as a mediator of the relationship of perceived situational factors to affective organisational commitment. *Applied Psychology: An International Review, 50,* 615-634.

Moorman, R. H., Blakely, G. L., & Niehoff, B. P. (1998). Does perceived organizational support mediate the relationship between procedural justice and organizational citizenship behavior? *Academy of Management Journal, 41,* 351-357.

Mowrey, M. E. (1998). *Inclusion, involvements, and commitment: Comparisons among full-time and part-time workers* (Unpublished doctoral dissertation). University of Iowa, Iowa City.

Mukherjee, K. (2010). Exploring the mediating effect of perceived organizational support between organizational justice dimensions and affective commitment. *Globsyn Management Journal, 4,* 9-22.

Muse, L. A., & Stamper, C. L. (2007). Perceived organizational support: Evidence for a mediated association with work performance. *Journal of Managerial Issues, 19,* 517-535.

Muse, L., Harris, S. G., Giles, W. F., & Feild, H. S. (2008). Work-life benefits and positive organizational behavior: Is there a connection? *Journal of Organizational Behavior, 29,* 171-192.

Ng, C., & Sarris, A. (2009). Distinguishing between the effect of perceived organisational support and person– organisation fit on work outcomes. *Australian & New Zealand Journal of Organisational Psychology, 2,* 1-9.

Nguyen, N. T. (2004). *The consequences of spatial distance and electronic communication for teleworkers: A multi-level investigation* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.

Nguyen, V. Q. (2011). *Organizational, job, and supervisory antecedents and consequence of job embeddedness: The case of Vietnam* (Unpublished doctoral dissertation). Mississippi State University, Starkville.

Nixon, A. E., Yang, L.-Q., Spector, P. E., & Zhang, X. (2011). Emotional labor in China: Do perceived organizational support and gender moderate the process? *Stress and Health, 27,* 289-305.

Norris, M. E. (2004). *Turnover in the military: Impact of workplace stressors* (Unpublished master's thesis). Saint Mary's University, Moraga, CA.

Nye, L. G., & Witt, L. A. (1993). Dimensionality and construct-validity of the Perceptions of Organizational Politics Scale (POPS). *Educational and Psychological Measurement, 53,* 821-829.

O'Brien, K. E., & Allen, T. D. (2008). The relative importance of correlates of organizational citizenship behavior and counterproductive work behavior using multiple sources of data. *Human Performance, 21,* 62-88.

O'Donohue, W., Donohue, R., & Grimmer, M. (2007). Research into the psychological contract: Two Australian perspectives. *Human Resource Development International, 10,* 301-318.

Pack, S. M. (2005). *Antecedents and consequences of perceived organizational support for NCAA athletic administrators* (Unpublished doctoral dissertation). Ohio State University, Columbus.

Paille, P., Bourdeau, L., & Galois, I. (2010). Support, trust, satisfaction, intent to leave and citizenship at organizational level: A social exchange approach. International Journal of *Organizational Analysis, 18,* 41-58.

Panaccio, A., & Vandenberghe, C. (2009). Perceived organizational support, organizational commitment and psychological well-being: A longitudinal study. *Journal of Vocational Behavior, 75,* 224-236.

Panatik, S. A. B. (2010). *Impact of work design on psychological work reactions and job performance among technical workers: A longitudinal study in Malaysia* (Unpublished doctoral dissertation). University of Waikato, Hamilton, New Zealand.

Pazy, A., & Ganzach, Y. (2009). Pay contingency and the effects of perceived organizational and supervisor support on performance and commitment. *Journal of Management, 35,* 1007-1025.

Peelle, H. E. (2007). Reciprocating perceived organizational support through citizenship behaviors. *Journal of Managerial Issues, 19,* 554-575.

Pernica, M. (2011). *Organizational change and employee trust: The mediating roles of perceived organizational support and organizational justice* (Unpublished master's thesis). Carleton University, Ottawa, Canada.

Perryer, C., Jordan, C., Firns, I., & Travaglione, A. (2010). Predicting turnover intentions: The interactive effects of organizational commitment and perceived organizational support. *Management Research Review, 33,* 911-923.

Phillips, G. M. (2000). *Perceived organizational support: An extended model of the mediating and moderating effects of self-structures* (Unpublished doctoral dissertation). University of Akron, Akron, OH.

Pianluprasidh, P. (2005). *The effect of perceived organizational support and organizational commitment on organizational citizenship behavior among nurses in Thailand* (Unpublished doctoral dissertation). Alliant International University, San Diego, CA.

Piasentin, K. A. (2007). *How do employees conceptualize fit? Development of a multidimensional measure of subjective person-organization fit* (Unpublished doctoral dissertation). University of Calgary, Calgary, Canada.

Piercy, N. F., Cravens, D. W., Lane, N., & Vorhies, D. W. (2006). Driving organizational citizenship behaviors and salesperson in-role behavior performance: The role of management control and perceived organizational support. *Journal of the Academy of Marketing Science, 34,* 244-262.

Polly, L. M. (2002). *Social exchange and customer service: The relationship between perceived organizational support, leader-member exchange, and customer service behavior* (Unpublished doctoral dissertation). Louisiana State University, Baton Rouge.

Potts, L. N., & Reynolds, D. (2010). The effect of supervisor-subordinate language similarity on subordinates' perceived organizational support in the U.S. lodging industry. *Journal of Human Resources in Hospitality & Tourism, 9,* 92-102.

Prehar, C. A. (2001). *Broadening our perspective of employee-organization linkages: From organizational commitment to organizational attachment* (Unpublished doctoral dissertation). Colorado State University, Fort Collins.

Purang, P. (2011). Organisational justice and affective commitment: The mediating role of perceived organisational support. *Asian Academy of Management Journal, 16,* 141-156.

Randall, D. M., & O'Driscoll, M. P. (1997). Affective versus calculative commitment: Human resource implications. *Journal of Social Psychology, 137,* 606-617.

Randall, M. L., Cropanzano, R., Bormann, C. A., & Birjulin, A. (1999). Organizational politics and organizational support as predictors of work attitudes, job performance, and organizational citizenship behavior. *Journal of Organizational Behavior, 20,* 159-174.

Reeves, J. N. (2011). *Engagement's role in the job characteristics model* (Unpublished doctoral dissertation). University of Houston, Houston, TX.

Reid, M. F., Riemenschneider, C. K., Allen, M. W., & Armstrong, D. J. (2008). Information technology employees in state government: A study of affective organizational commitment, job involvement, and job satisfaction. *The American Review of Public Administration, 38,* 41-61.

Reinardy, S. (2006). *Beyond the games: A study of the effects of life issues and burnout on newspaper sports editors* (Unpublished doctoral dissertation). University of Missouri–Columbia.

Reinardy, S. (2009). Beyond satisfaction: Journalists doubt career intentions as organizational support diminishes and job satisfaction declines. *Atlantic Journal of Communication, 17,* 126-139.

Rever-Moriyama, S. D. (1996). *Antecedents of organizational citizenship and job search behaviours in university professors* (Unpublished master's thesis). University of Calgary, Calgary, Canada.

Reynolds, K. D. (2009). *The effects of abusive supervision and social support on workplace aggression* (Unpublished master's thesis). DePaul University, Chicago, IL.

Rhee, J., Park, T., & Hwang, S. H. (2011). Non-regular professionals' dual commitment in South Korea: Antecedents and consequences. *The International Journal of Human Resource Management, 22,* 612-631.

Rhoades, L., Eisenberger, R., & Armeli, S. (2001). Affective commitment to the organization: The contribution of perceived organizational support. *Journal of Applied Psychology, 86,* 825-836.

Rich, B. L., Lepine, J. A., & Crawford, E. R. (2010). Job engagement: Antecedents and effects on job performance. *Academy of Management Journal, 53,* 617-635.

Richter, L. K. (2001). *Factors affecting exchange relationships among subordinates and supervisors: A study of military officers* (Unpublished master's thesis). Air Force Institute of Technology, Wright-Patterson AFB, OH.

Riddel, K. C. (2010). *An analysis of factors that influence air force civil engineer company grade officer turnover intentions* (Unpublished master's thesis). Air Force Institute of Technology, Wright-Patterson AFB, OH.

Riggle, R. J. (2008). *The impact of organizational climate variables of perceived organizational support, workplace isolation, and ethical climate on salesperson psychological and behavioral work outcomes* (Unpublished doctoral dissertation). University of South Florida, Tampa.

Ristig, K. (2004). *Antecedents and consequences of trust within organizations* (Unpublished doctoral dissertation). Louisiana Tech University, Ruston, TX.

Roch, S. G., & Shanock, L. R. (2006). Organizational justice in an exchange framework: Clarifying organizational justice distinctions. *Journal of Management, 32,* 299-322.

Ruggiere, P. J. (1991). *Organizational commitment in a self-managing work team environment* (Unpublished doctoral dissertation). University of North Texas, Denton, TX.

Rutherford, B., Boles, J. S., Hamwi, G. A., & Rutherford, L. G. (2010). Perceived organizational support and the seven facets of salesperson job satisfaction. *Journal of Selling & Major Account Management, 10,* 8-20.

Rutherford, B., Park, J., & Han, S.-L. (2011). Increasing job performance and decreasing salesperson propensity to leave: An examination of an Asian sales force. *Journal of Personal Selling & Sales Management, 31,* 171-183.

Rutherford, M. W., & Holt, D. T. (2007). Corporate entrepreneurship: An empirical look at the innovativeness dimension and its antecedents*. Journal of Organizational Change Management, 20,* 429-446.

Salim, L. (2005). *Organizational commitment in a health NGO in Pakistan* (Unpublished doctoral dissertation). University of Toronto, Toronto, Canada.

Sarabia, C. L. (2002). *An examination of the relationship between career development and organizational commitment from a human capital perspective* (Unpublished doctoral dissertation). University of Georgia, Athens, GA.

Scher, S. L. (1991). *Psychologists in state psychiatric settings: Attributional style, self-efficacy, and perceived organizational support as correlated of job satisfaction and burnout* (Unpublished doctoral dissertation). Hofstra University, Hempstead, NY.

Self, D. R., Holt, D. T., & Schaninger, W. S. (2005). Work-group and organizational support: A test of distinct dimensions. *Journal of Occupational and Organizational Psychology, 78,* 133-140.

Settoon, R. P., Bennett, N., & Liden, R. C. (1996). Social exchange in organizations: Perceived organizational support, leader-member exchange, and employee reciprocity. *Journal of Applied Psychology, 81,* 219-227.

Shaffer, M. A., Harrison, D. A., Gilley, K. M., & Luk, D. M. (2001). Struggling for balance amid turbulence on international assignments: Work-family conflict, support and commitment. *Journal of Management, 27,* 99-121.

Shalhoop, J. H. (2003). *Social-exchange as a mediator of the relationship between organizational justice and workplace outcomes* (Unpublished doctoral dissertation). University of Akron, Akron, OH.

Shanock, L. R., & Eisenberger, R. (2006). When supervisors feel supported: Relationships with subordinates' perceived supervisor support, perceived organizational support, and performance. *Journal of Applied Psychology, 91,* 689-695.

Sherman, M. P. (2004). *Antecedents of affective commitment among subgroups of part-time workers: All parttimers are not created equal* (Unpublished doctoral dissertation). Wayne State University, Detroit, MI.

Shore, L. M., & Barksdale, K. (1998). Examining degree of balance and level of obligation in the employment relationship: A social exchange approach. *Journal of Organizational Behavior, 19,* 731-744.

Shore, L. M., & Tetrick, L. E. (1991). A construct validity study of the Survey of Perceived Organizational Support. *Journal of Applied Psychology, 76,* 637-643.

Shore, L. M., & Wayne, S. J. (1993). Commitment and employee behavior: Comparison of affective commitment and continuance commitment with perceived organizational support. *Journal of Applied Psychology, 78,* 774-780.

Shore, L. M., Tetrick, L. E., Lynch, P., & Barksdale, K. (2006). Social and economic exchange: Construct development and validation. *Journal of Applied Social Psychology, 36,* 837-867.

Showail, S. J. (2007). *Foreign workers in Saudi Arabia and their sources of adjustment: The role of supervisor and spouse report* (Unpublished doctoral dissertation). Washington University in St. Louis, St. Louis, MO.

Slan Jerusalim, R. S. (2009). *The impact of succession management systems on manager attitudes and organizational outcomes* (Unpublished doctoral dissertation). University of Guelph, Guelph, Canada.

Smith, M. A. (2002). *Reciprocity and social exchange relationships in organizations: Examining why and how individuals contribute to organizational social capital* (Unpublished doctoral dissertation). University of Pittsburgh, Pittsburgh, PA.

Snape, E., & Redman, T. (2010). HRM practices, organizational citizenship behaviour, and performance: Amulti-level analysis. *Journal of Management Studies, 47,* 1219-1247.

Spitzmu¨ ller, C., Glenn, D. M., Barr, C. D., Rogelberg, S. G., & Daniel, P. (2006). ‘‘If you treat me right, I reciprocate’’: Examining the role of exchange in organizational survey response. *Journal of Organizational Behavior, 27,* 19-35.

Stamper, C. L., & Johlke, M. C. (2003). The impact of perceived organizational support on the relationship between boundary spanner role stress and work outcomes. *Journal of Management, 29,* 569-588.

Stinglhamber, F., & Vandenberghe, C. (2003). Organizations and supervisors as sources of support and targets of commitment. *Journal of Organizational Behavior, 24,* 251-270.

Stinglhamber, F., De Cremer, D., & Mercken, L. (2006). Perceived support as a mediator of the relationship between justice and trust: A multiple foci approach. *Group and Organization Management, 31,* 442-468.

Stru¨mpfer, D. W., Hardy, A., de Villiers, J. S., & Rigby, S. (2009). Organisationally relevant variables and Keyes's mental health continuum scales: An exploratory study. South African *Journal of Industrial Psychology, 35,* 165-171.

Stup, R. E. (2006). *Human resource management, organizational commitment, and perceived organizational support in dairy farm businesses* (Unpublished doctoral dissertation). Pennsylvania State University, University Park, PA.

Suazo, M. M. (2003). *An examination of antecedents and consequences of psychological contract breach* (Unpublished doctoral dissertation). University of Kansas, Manhattan, KS.

Suazo, M. M. (2009). The mediating role of psychological violation on the relations between psychological contract breach and work-related attitudes and behaviors. *Journal of Managerial Psychology, 24,* 136-60.

Sucharski, I. L. (2007). *Influencing employees' generalization of support and commitment from supervisor to organization* (Unpublished doctoral dissertation). University of Delaware, Newark.

Suckow, K. J. (1999). *The effects of dual employment on temporary employees* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.

Sung, E. C. (1999). *Effects of an ergonomic intervention on psychosocial outcomes* (Unpublished master's thesis). San Jose State University, San Jose, CA.

Susa, A. M. (2003). *Humor type, organizational climate, and outcomes: The shortest distance between an organization's environment and the bottom line is laughter* (Unpublished doctoral dissertation). University of Nebraska–Lincoln, NB

Susskind, A. M., Borchgrevink, C. P., Kacmar, K. M., & Brymer, R. A. (2000). Customer service employees'behavioral intentions and attitudes: An examination of construct validity and a path model. *International Journal of Hospitality Management, 19,* 53-77.

Tan, F. M. (2008). Organisational support as the mediator of career-related HRM practices and affective commitment: Evidence from knowledge workers in Malaysia. Research and Practice in *Human Resource Management, 16,* 8-24.

Taveras, L. E. (2000). *The effects of telecommuting on attitudinal and behavioral characteristics of organizational culture* (Unpublished doctoral dissertation). University of Sarasota, Sarasota, FL.

Tekleab, A. G., & Chiaburu, D. S. (2011). Social exchange: Empirical examination of form and focus. *Journal of Business Research, 64,* 460-466.

Tekleab, A. G., Takeuchi, R., & Taylor, M. S. (2005). Extending the chain of relationships among organizational justice, social exchange, and employee reactions: The role of contract violations. *Academy of Management Journal, 48,* 146-157.

Tolentino, A. L. (2009). *Are all good soldiers created equal? Examining the ''why'' that underlies organizational citizenship behavior: The development of an OCB motives scale* (Unpublished doctoral dissertation). University of South Florida, Tampa, FL.

Tortella, J. A. (2009). *Operations tempo and turnover intentions: An exploratory study of the Air Force's explosive ordnance disposal (EOD) career field and development of the Air Force civil engineer retention questionnaire* (Unpublished master's thesis). Air Force Institute of Technology, Wright-Patterson AFB, OH.

Tourangeau, A., Cranley, L., Spence Laschinger, H. K., & Pachis, J. (2010). Relationships among leadership practices, work environments, staff communication and outcomes in long-term care. *Journal of Nursing Management, 18,* 1060-1072.

Treadway, D. C., Hochwarter, W. A., Ferris, G. R., Kacmar, C. J., Douglas, C., Ammeter, A. P., & Buckley, M. R. (2004). Leader political skill and employee reactions. *Leadership Quarterly, 15,* 493-513.

Tumwesigye, G. (2010). The relationship between perceived organizational support and turnover intentions in a developing country: The mediating role of organizational commitment. African *Journal of Business Management, 46,* 942-952.

Uhl-Bien, M., & Maslyn, J. M. (2003). Reciprocity in manager-subordinate relationships: Components, configurations, and outcomes. *Journal of Management, 29,* 511-532.

Valentine, S., Greller, M. M., & Richtermeyer, S. B. (2006). Employee job response as a function of ethical context and perceived organizational support. *Journal of Business Research, 59,* 582-588.

Valentino, K. E. M. (1995). *The effects of leadership style and behavior on perceived organizational support and organizational commitment* (Unpublished master's thesis). California State University, Long Beach, CA.

Van Dam, K. (2004). Antecedents and consequences of employability orientation. *European Journal of Work and Organizational Psychology, 13,* 29-51.

Van Knippenberg, D., & Sleebos, E. (2006). Organizational identification versus organizational commitment: Self-definition, social exchange, and job attitudes. *Journal of Organizational Behavior, 27,* 571-584.

Vandenberghe, C., & Peiro, J. (1999). Organizational and individual values: Their main and combined effects on work attitudes and perceptions. *European Journal of Work & Organizational Psychology, 8,* 569-581.

Vandenberghe, C., Bentein, K., & Stinglhamber, F. (2004). Affective commitment to the organization, supervisor, and work group: Antecedents and outcomes. *Journal of Vocational Behavior, 64,* 47-71.

Vandenberghe, C., Bentein, K., Michon, R., Chebat, J-C., Tremblay, M., & Fils, J-F. (2007). An examination of the role of perceived support and employee commitment in employee-customer encounters. *Journal of Applied Psychology, 92,* 1177-1187.

Venkatachalam, M. (1995). *Personal hardiness and perceived organizational support as links in the role stressoutcome relationship: A person-environment fit model* (Unpublished doctoral dissertation). University of Alabama, Tuscaloosa, AL.

Villanueva, D., & Djurkovic, N. (2009). Occupational stress and intention to leave among employees in small and medium enterprises. *International Journal of Stress Management, 16,* 124-137.

Wahab, E. (2010, June). *The moderating role of power distance on the relationship between perceived organizational support and organizational commitment.* Paper presented at the 5th National Human Resource Management Conference, Kuala Terengganu, Malaysia.

Wang, G., & Lee, P. D. (2009). Psychological empowerment and job satisfaction. Group & *Organizational Management, 34,* 271-296.

Wang, G., Liu, D., & Wang, X. (2011). Effects of perceived organizational support and Guanxi on salesperson performance: The mediation of customer need knowledge. *Frontiers of Business Research in China, 5,* 422-435.

Wang, M-L. (2009a). What makes a good citizen in service settings? *The Service Industries Journal, 29,* 621-634.

Wang, M., & Takeuchi, R. (2007). The role of goal orientation during expatriation: A cross-sectional and longitudinal investigation. *Journal of Applied Psychology, 92,* 1437-1445.

Watt, J. D., & Hargis, M. B. (2010). Boredom proneness: Its relationship with subjective underemployment, perceived organizational support, and job performance. *Journal of Business and Psychology, 25,* 163-174.

Waul, A. R. (2007). *An empirical study of leader member exchange (LMX) on consultants' attitudes and behaviors towards their parent organization* (Unpublished doctoral dissertation). Touro University International, Cypress, CA.

Wayne, S. J., Shore, L. M., & Liden, R. C. (1997). Perceived organizational support and leader-member exchange: A social exchange perspective. *Academy of Management Journal, 40,* 82-111.

Wayne, S. J., Shore, L. M., Bommer, W. H., & Tetrick, L. E. (2002). The role of fair treatment and rewards in perceptions of organizational support and leader-member exchange. *Journal of Applied Psychology, 87,* 590-598.

Webster, J. R., & Adams, G. A. (2010). Organizational support and contract fulfillment as moderators of the relationship between preferred work status and performance. *Journal of Business and Psychology, 25,* 131-138.

Welchans, T. D. (1996). *The effects of telecommuting and communication media on perceived value congruence, organizational support and job satisfaction* (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

Wettlaufer, A. (2009). *Determinants and moderators of burnout and job satisfaction among managers of volunteers in the community support sector* (Unpublished master's thesis). University of Guelph, Guelph, Canada.

Williams, M. L., Brower, H. H., Ford, L. R., Williams, L. J., & Carraher, S. M. (2008). A comprehensive model and measure of compensation satisfaction. *Journal of Occupational and Organizational Psychology, 81,* 639-668.

Wilson, K. T. (2007). *Technology usefulness and impact on school foodservice employees' perceptions of organizational support* (Unpublished doctoral dissertation). Iowa State University, Ames.

Wilson, M. G., Dejoy, D. M., Vandenberg, R. J., Richardson, H. A., & McGrath, A. L. (2004). Work characteristics and employee health and well-being: Test of a model of healthy work organization. *Journal of Occupational and Organizational Psychology, 77,* 565-588.

Witt, L. A. (1992). Exchange ideology as a moderator of the relationships between importance of participation in decision making and job attitudes. *Human Relations, 45,* 73-85.

Witt, L. A. (1994). *Perceptions of organizational support and affectivity as predictors of job satisfaction* (DOT/FAA/AM-94-2). Washington, DC: Office of Aviation Medicine.

Witt, L. A., & Carlson, D. S. (2006). The work-family interface and job performance: Moderating effects of conscientiousness and perceived organizational support. Journal of *Occupational Health Psychology, 11,* 343-357.

Witt, L. A., & Nye, L. G. (1992). *Organizational goal congruence and job attitudes revisited* (FAA-AM Publication No. 92-8, pp. 92-98). Washington, DC: Federal Aviation Administration, Office of Aviation Medicine.

Woodcock, C. (2010). *Transformational leadership and employee commitment: The moderating role of perceived organizational support and individual level collectivism/individualism* (Unpublished master's thesis). Northern Illinois University, DeKalb, IL.

Woodward, C. (2006). *Correlates of re-enlistment intentions among army National Guard members* (Unpublished doctoral dissertation). San Jose State University, San Jose, CA.

Worley, J. A., Fuqua, D. R., & Hellman, C. M. (2009). The survey of perceived organisational support: Which measure should we use? *South African Journal of Industrial Psychology, 35,* 112-116.

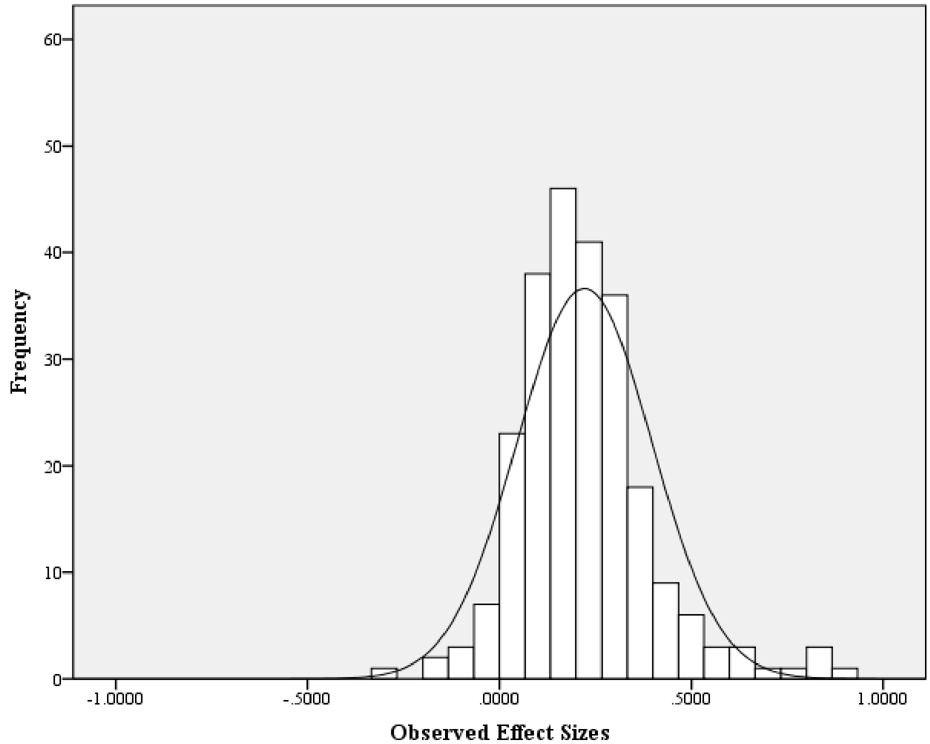
Yoon, J., & Thye, S. R. (2002). A dual process model of organizational commitment: Job satisfaction and organizational support. *Work and Occupations, 29,* 97-124.

Yoshimura, K. E. (2003). *Employee traits, perceived organizational support, supervisory communication, affective commitment, and intent to leave: Group differences* (Unpublished doctoral dissertation). North Carolina State University, Raleigh, NC.

APPENDIX B

OVERALL DISTRIBUTION OF THE OBSERVED EFFECT SIZES

Overall distribution of the observed effect sizes for the correlation of POS with its criterion variables.



APPENDIX C

IRB

Protection of Human Rights

The current meta-analysis research involves the collection of existing data, documents, and records that are publicly available. Information is recorded at the study level by the investigator in a manner that subjects from each study cannot be identified, directly or through identifiers linked to the subjects. Therefore, the current study meets the IRB exemption criteria listed in 45 CFR 46.101(b) because no human subjects will be directly involved in the current study.

REFERENCE LIST

Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011). Meta-Analytic Choices and Judgment Calls: Implications for Theory Building and Testing, Obtained Effect Sizes, and Scholarly Impact. *Journal of Management, 37*(1), 5-38.

Alexander, R. A., Carson, K. P., Alliger, G. M., & Cronshaw, S. F. (1989). Empirical Distributions of Range Restricted SD x in Validity Studies. *Journal of Applied Psychology, 74*(2), 253.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, Calif.: Brooks/Cole Pub.

Aytug, Z., Rothstein, H., Zhou, W., & Kern, M. (2012). Revealed or Concealed? Transparency of Procedures, Decisions, and Judgment Calls in Meta-Analyses. *Organizational Research Methods, 15*(1), 103-133.

Behrens, John T. (1997). Does the White Racial Identity Attitude Scale Measure Racial Identity? *Journal of Counseling Psychology, 44*(1), 3-12.

Beretvas, S., & Pastor, D. (2003). Using mixed-effects models in reliability generalization studies. *Educational and Psychological Measurement, 63*(1), 75-95.

Bishara, A., & Hittner, J. (2012). Testing the Significance of a Correlation with Nonnormal Data: Comparison of Pearson, Spearman, Transformation, and Resampling Approaches. *Psychological Methods, 17*(3), 399-417.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: Wiley.

Bonett, D. (2010). Varying coefficient meta-analytic methods for alpha reliability. *Psychological Methods, 15*(4), 368-85.

Borman, W., & Motowidlo, S. (1997). Task Performance and Contextual Performance: The Meaning for Personnel Selection Research. *Human Performance, 10*(2), 99-109.

Brannick, M. T., & Zhang, N. (2013). Bayesian meta-analysis of coefficient alpha. *Research Synthesis Methods, 4,* 2, 198-207.

Callander, J.C., Osburn, H.G., Greener, J.M., & Ashworth, S. (1982). Multiplicative Validity Generalization Model: Accuracy of Estimates as a Function of Sample Size and Mean; Variance and Shape of the Distribution of True Validities. *Journal of Applied Psychology, 67*, 859-867.

Callender, J. C., & Osburn, H. G. (1980). Development and test of a new model for validity generalization. *Journal of Applied Psychology, 65,* 543–558.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), New Jersey: Lawrence Erlbaum Associates

Cooper, H. (1998). *Integrating research: A guide for literature reviews* (3rd ed.). Newbury Park, CA: Sage.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston.

DeSimone, Justin A. (2014). When It's Incorrect to Correct: A Brief History and Cautionary Note. *Industrial & Organizational Psychology, 7*(4), 527-532.

Eisenberger, R., Cummings, J., Armeli, S., & Lynch, P. (1997). Perceived organizational support, discretionary treatment, and job satisfaction. *Journal of Applied Psychology, 82,* 812–820.

Eisenberger, R., Fasolo, P., Davis-LaMastro, V., & Schmitt, Neal. (1990). Perceived Organizational Support and Employee Diligence, Commitment, and Innovation. *Journal of Applied Psychology, 75*(1), 51-59.

Eisenberger, R., Huntington, R., Hutchison, S., Sowa, D., & Guion, Robert. (1986). Perceived Organizational Support. *Journal of Applied Psychology, 71*(3), 500-507.

Feldt, L. S. (1965). The approximate sampling distribution of Kuder–Richardson reliability coefficient twenty. *Psychometrika, 30,* 357–370.

Field, A. (2001). Meta-Analysis of Correlation Coefficients: A Monte Carlo Comparison of Fixed- and Random-Effects Methods. *Psychological Methods, 6*(2), 161-180.

Graham, J. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability - What they are and how to use them. *Educational and Psychological Measurement, 66*(6), 930-944.

Hall, S. M., & Brannick, M. T. (2002). Comparison of Two Random-Effects Methods of Meta-Analysis. *Journal of Applied Psychology, 87*(2), 377-389.

Hedges, L. & Vevea, J. (1998). Fixed- and Random-Effects Models in Meta-Analysis. *Psychological Methods, 3*(4), 486-504.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in meta-analysis. *Statistics in Medicine, 21,* 1539–1558.

Huedo-Medina, T., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11*(2), 193-206.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

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

Hunter, J. E., Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91,* 594–612.

Köhler, T., Cortina, J., Kurtessis, J., & Gölz, M. (2015). Are We Correcting Correctly? Interdependence of Reliabilities in Meta-Analysis. *Organizational Research Methods, 18*(3), 355-428.

Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika, 28,* 221–238.

Kurtessis, J. N., Eisenberger, R. Ford, M. T. Buffardi, L. C. Stewart, K. A., & Adis, C. S. (2015). Perceived organizational support: A meta-analytic evaluation of organizational support theory. *Journal of Management,* .1-31. DOI: 10.1177/0149206315575554

Law, K., Schmidt, F., Hunter, J., & Schmitt, Neal. (1994). A Test of Two Refinements in Procedures for Meta-Analysis. *Journal of Applied Psychology, 79*(6), 978-986.

Lambert, R. G., & Curlette, W. L. (1995, April). *The Robustness of the Standard Error of Summarized, Corrected Validity Coefficients to Non-Independence and Non-Normality of Primary Data.* Paper presented at the Annual AERA Conference, San Francisco, CA.

Le, H., & Schmidt, F. L. (2006). Correcting for indirect range restriction in meta-analysis: testing a new meta-analytic procedure. *Psychological Methods, 11,* 4, 416-38.

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, 7,* 4, 478-500.

Li, Johnson C. H. (2013.). *Evaluating the Performance of the Uncorrected and Corrected Reliability Alpha for Range Restriction and the Confidence Intervals in a Single and Meta-Analytic Study* (Unpublished doctoral dissertation). University of Alberta, Alberta, Canada.

Lord, F. M. (1974). Variance stabilizing transformation of the stepped-up reliability coefficient. *Journal of Educational Measurement, 11,* 55–57.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Pub. Co.

McNemar, Q. (1969). *Psychological statistics*. New York: Wiley.

Mendoza, J. L., & Mumford, M. (1987). Correction for attenuation and range restriction on the predictor. *Journal of Educational Statistics, 12,*282–293.

Mendoza, J., & Reinhardt, R. (1991). Validity Generalization Procedures Using Sample-Based Estimates: A Comparison of Six Procedures. *Psychological Bulletin, 110*(3), 596.

Moorman, R., & Blakely, G. (1995). Individualism collectivism as an individual difference predictor of organizational citizenship behavior. *Journal of Organizational Behavior (1986-1998), 16*(2), 127.

Murphy, K. R. (2003). *Validity generalization: A critical review*. Mahwah, NJ: Lawrence Erlbaum Associates.

Murphy, K., & Cleveland, J. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Thousand Oaks, Calif.: Sage Publications.

Murphy, K., & DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology, 53*(4), 873-900.

Ones, D., Viswesvaran, C., & Schmidt, F. (2008). No New Terrain: Reliability and Construct Validity of Job Performance Ratings. *Industrial and Organizational Psychology, 1*(2), 174-179.

Organ, D.W. (1988). *Organizational citizenship behavior: The good soldier syndrome.* Lexington, MA: D.C. Heath.

Oswald, F. L., & Johnson, J. W. (1998). On the robustness, bias, and stability of statistics from meta-analysis of correlation coefficients: Some initial Monte Carlo findings. *Journal of Applied Psychology, 83(2),* 164-178.

Paese, P., Switzer, F., & Guion, Robert. (1988). Validity Generalization and Hypothetical Reliability Distributions: A Test of the Schmidt–Hunter Procedure. *Journal of Applied Psychology, 73*(2), 267-274.

Pearlman, K., Schmidt, F., & Hunter, J. (1980). Validity Generalization Results for Tests Used to Predict Job Proficiency and Training Success in Clerical Occupations. *Journal of Applied Psychology, 65*(4), 373.

Podsakoff, P. M., & MacKenzie, S. B. (1989). *A Second Generation Measure of Organizational Citizenship Behavior* (Unpublished manuscript). Indiana University, Bloomington.

Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity generalization. *Journal of Applied Psychology, 68,* 382–395.

Raju, N., & Brand, P. (2003). Determining the significance of correlations corrected for unreliability and range restriction. *Applied Psychological Measurement, 27*(1), 52-71.

Raju, N., Anselmi, T., Goodman, J., & Thomas, A. (1998). The effect of correlated artifacts and true validity on the accuracy of parameter estimation in validity generalization. *Personnel Psychology, 51*(2), 453-465.

Raju, N., Burke, M., Normand, J., & Langlois, G. (1991). A new meta-analytic approach. *Journal of Applied Psychology, 76*(3), 432-446.

Raju, N., Pappas, S., Williams, C., & Schmitt, Neal. (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*(6), 901-911.

Rhoades, L., & Eisenberger, R. (2002). Perceived organizational support: A review of the literature. *Journal of Applied Psychology, 87,* 698-714.

Sawilowsky, Shlomo S. (2000). Psychometrics Versus Data metrics: Comment on Vacha-Haase's "Reliability Generalization" Method and Some "EPM" Editorial Policies. *Educational and Psychological Measurement, 60*(2), 157-173.

Schmidt, F. & Hunter, J. (1996). Measurement Error in Psychological Research: Lessons from 26 Research Scenarios. *Psychological Methods, 1*(2), 199-223.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62,* 529-540.

Schmidt, F. L., Gast-Rosenberg, I. F., & Hunter, J. E. (1980). Validity generalization results for computer programmers. *Journal of Applied Psychology, 65,* 643–661.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976) Statistical power in criterion-related validity studies. *Journal of Applied Psychology, 61,* 473-485.

Schmidt, F. L., Law, K., Hunter, J.E., Rothstein, H. R., Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. *Journal of Applied Psychology, 78,* 3-12.

Schmidt, F., Le, H., & Ilies, R. (2003). Beyond Alpha: An Empirical Examination of the Effects of Different Sources of Measurement Error on Reliability Estimates for Measures of Individual Differences Constructs.*Psychological Methods, 8*(2), 206-224.

Schmidt, F., Viswesvaran, C., & Ones, D. (2000). Reliability is not validity and validity is not reliability. *Personnel Psychology, 53*(4), 901-912.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*(1), 6.

Van Dyne, L., Graham, J., & Dienesch, R. (1994). Organizational citizenship behavior: Construct redefinition, measurement, and validation. (Includes appendix). *Academy of Management Journal, 37*(4), 765.

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics, 30*(3), 261-293.

Viechtbauer, W. (2010). Conducting meta-analysis in R with the Metafor package. *Journal of Statistical Software, 36,* 1–48.

Viechtbauer, W. (2016). The Metafor Package: A Meta-Analysis Package for R. Retrieved from http://www.metafor-project.org/doku.php/tips:hunter_schmidt_method?s[]=hunter&s[]=schmidt

Williams, L., & Anderson, S. (1991). Job Satisfaction and Organizational Commitment as Predictors of Organizational Citizenship and In-Role Behaviors. *Journal of Management, 17*(3), 601.

Worthen, B. R., White, K. R., Fan, X., & Sudweeks, R. R. (1999). *Measurement and assessment in the schools* (2nd ed.). New York: Longman.

Zimmerman, D., & Williams, R. (1977). The theory of test validity and correlated errors of measurement. *Journal of Mathematical Psychology, 16,* 135–152.

VITA

Lei (Kelly) Zhao graduated from Illinois Institute of Technology (IIT) in 2010 with a Master of Science degree in Personnel and Human Resources Development. She was admitted to the Ph.D. program of Research and Methodology at Loyola University Chicago in August 2011. Throughout her time at school, she was funded by IIT's Department of Psychology and Loyola's School of Education. Dr. Zhao had also expanded her academic trainings to Biostatistics and Statistical Design of Experiments during her time at Loyola. She has been working at the John G. Shedd Aquarium for more than six years as a data analyst. Her academic interests are in research design, meta-analysis, psychometrics in Industrial and Organizational Psychology, Market Research, and Biostatistics.