Meta-Analytic Choices and Judgment Calls: Implications for Theory Building and Testing, Obtained Effect Sizes, and Scholarly Impact

Herman Aguinis
Dan R. Dalton
Indiana University
Frank A. Bosco
Charles A. Pierce
University of Memphis
Catherine M. Dalton
Indiana University

The authors content analyzed 196 meta-analyses including 5,581 effect-size estimates published in Academy of Management Journal, Journal of Applied Psychology, Journal of Management, Personnel Psychology, *and* Strategic Management Journal *from January 1982 through August 2009 to assess the presumed effects of each of 21 methodological choices and judgment calls on substantive conclusions. Results indicate that, overall, the various meta-analytic methodological choices available and judgment calls involved in the conduct of a meta-analysis have little impact on the resulting magnitude of the meta-analytically derived effect sizes. Thus, the present study, based on actual meta-analyses, casts doubt on previous warnings, primarily based on selective case studies, that judgment calls have an important impact on substantive conclusions. The authors also tested the fit of a multivariate model that includes relationships among*

5

theory-building and theory-testing goals, obtained effect sizes, year of publication of the meta-analysis, and scholarly impact (i.e., citations per year). Results indicate that the more a meta-analysis attempts to test an existing theory, the larger the number of citations, whereas the more a meta-analysis attempts to build new theory, the lower the number of citations. Also, in support of scientific particularism, as opposed to scientific universalism, the magnitude of the derived effects is not related to the extent to which a meta-analysis is cited. Taken together, the results provide a comprehensive data-based understanding of how meta-analytic reviews are conducted and the implications of these practices for theory building and testing, obtained effect sizes, and scholarly impact.

Meta-analyses (i.e., quantitative literature reviews) have revolutionized the field of management. Conclusions based on meta-analytic findings set the standard for what is considered state of the science, for what we know and do not know, and for which theory is considered valid and which one is not (Schmidt, 1992, 1996). Not surprisingly, as we describe in a subsequent section herein, articles reporting meta-analyses are cited at a significantly higher rate than are primary-level studies.

Dozens of articles addressing the choices that meta-analysts face have been published on a regular basis since the early 1980s (e.g., Aguinis, 2001; Aguinis, Gottfredson, & Wright, in press; Aguinis & Whitehead, 1997; Callender & Osburn,1988; Erez, Bloom, & Wells, 1996; Geyskens, Krishnan, Steenkamp, & Cunha, 2009; Hunter, Schmidt, & Le, 2006; James, Demaree, & Mulaik, 1986; James, Demaree, Mulaik, & Ladd, 1992; Law, Schmidt, & Hunter, 1994; McDaniel, Rothstein, & Whetzel, 2006; Newman, Jacobs, & Bartram, 2007; Osburn & Callender, 1992; Oswald & Johnson, 1998; Schmidt & Raju, 2007; Schmidt, Shaffer, & Oh, 2008; Steel & Kammeyer-Mueller, 2002). In addition to regular articles, journals have published special issues devoted to meta-analytic methodology, such as a recent one in *Organizational Research Methods* (e.g., Aguinis, Sturman, & Pierce, 2008; Bobko & Roth, 2008; Dalton & Dalton, 2008; Kisamore & Brannick, 2008; Schmidt, 2008). Many of the articles on meta-analytic procedures have received an immense amount of attention. In fact, Murphy and Newman (2003) noted that "authors in other areas . . . would love to get the sort of attention the most pedestrian VG [validity generalization] paper was likely to receive" (p. 418). This large body of research compares various types of methodological choices faced by meta-analysts. These choices range from whether and how to eliminate certain primary-level studies from a meta-analytic database, the type of meta-analytic model to use, whether and how a meta-analysis should correct for various methodological and statistical artifacts, and whether and how to compute a fail-safe N statistic (e.g., the number of statistically nonsignificant primary-level studies that would be necessary to add to a meta-analysis to reduce a meta-analytically derived summary effect size to a statistically nonsignificant value; Rosenthal, 1979).

The large number of articles on meta-analytic procedures over the past quarter of a century leads to the conclusion that researchers interested in conducting a meta-analysis are faced with many methodological choices (Kulinskaya, Morgenthaler, & Staudte, 2008; Schulze,

Holling, & Bohning, 2003). Moreover, these methodological choices presumably have important implications in terms of the meta-analytic results and subsequent implications for theory and applications (e.g., Aguinis & Pierce, 1998; Johnson, Mullen, & Salas, 1995; Schmidt, 2008). For example, Oh, Schmidt, Shaffer, and Le (2008) used an improved range restriction correction and concluded that "the validity of the GMAT [Graduate Management Admission Test] has been underestimated by 7% due to the application of suboptimal range restriction corrections" (p. 563). In other words, it seems that how researchers make meta-analytic choices presumably affects substantive results in terms of the obtained effect-size estimates. In turn, these effect-size estimates are used to make decisions about the relative validity of a theory (e.g., Is personality related to job performance?) and the practical usefulness of an intervention (e.g., Is it equivalent to use subjective or objective measures for appraising employee performance?). In short, each one of numerous meta-analytic choices presumably has a direct impact on substantive theory-related conclusions as well as on practical applications (Schulze et al., 2003). This conclusion has been reached mostly via Monte Carlo simulation studies. But, is this conclusion true in actual meta-analytic practice? Have these methodological choices affected the obtained effect sizes in important ways and, thus, improved the predictive potential and usefulness of our theories?

In addition to the large body of research regarding various methodological choices available to meta-analysts, several articles have been published regarding judgment calls that meta-analysts often make but seldom acknowledge explicitly. For example, Bullock and Svyantek (1985) noted that researchers are usually not held accountable for whether their meta-analysis "avoids selecting studies based on criteria of methodological rigor, age of study, or publication status" (p. 114). Similarly, Wanous, Sullivan, and Malinak (1989) issued the following warning:

> It is ironic that, although meta-analysts have been known to criticize reporters of primary research for failing to report information crucial to a meta-analysis, meta-analysts themselves are guilty of a similar sin of omission, that is, the failure to report judgments made and their effects. (p. 263)

Several other authors have echoed the sentiment that judgment calls play an important role in affecting meta-analytically derived effect-size estimates. For example, Beal, Cohen, Burke, and McLendon (2003) asserted that "perhaps one of the largest sources of error in meta-analyses is the multitude of judgment calls made at various stages of the research synthesis process" (p. 993), and Colquitt, LePine, and Noe (2000) referred to the "subjectivity and judgment calls inherent in meta-analytic efforts" (p. 685). But are these warnings justified? What are the effects of these and other judgment calls on meta-analytic results? Have these seemingly idiosyncratic decisions had an impact on substantive conclusions?

Present Study

Our study examines empirically the extent to which methodological choices and judgment calls have affected meta-analytically obtained effect sizes and hence substantive conclusions

regarding the predictive potential and usefulness of theories. To do so, we conducted a content analysis of 196 meta-analyses published in the *Academy of Management Journal* (*AMJ*), *Journal of Applied Psychology* (*JAP*), *Journal of Management* (*JOM*), *Personnel Psychology* (*PPsych*), and *Strategic Management Journal* (*SMJ*) from January 1982 through August 2009. Taken together, these studies include a total of 5,581 meta-analytically derived effect sizes.

The specific goals of our study are as follows. First, we describe the typical set of methodological choices and judgment calls that authors have made. This description allows us to uncover the accepted norms, many of them implicit, regarding how to conduct a meta-analytic review.

Second, we assess the extent to which each of 21 meta-analytic methodological choices and judgment calls is related to obtained effect sizes. The practical usefulness of a theory is, in part, based on the extent to which it is able to explain and predict phenomena of interest (Bierstedt, 1959). Regardless of whether one predicts a positive or negative relationship between variables, the larger the effect size expressing this relationship in absolute value, the greater the predictive potential, and the better the usefulness of a theory. This is a fundamental principle in philosophy of science that dates back to the ideas set forth by John Stuart Mill (1843/2002). For example, an absolute effect size $|r| = .4$ indicates that a theory is better at predicting outcomes compared with an effect size $|r| = .1$ because in the former case 16% of variance is explained whereas in the latter case only 1% of the variance is explained. In some cases, it may be relevant to know that a variable is not related to a specific outcome, or a researcher may predict the presence of a small effect, for example that $\rho = .10$. If a researcher conducts a study and finds that $r = .10$, then he or she is correct in terms of the size of the predicted relationship. However, although the researcher is correct, his or her theory about the relationship may not be very useful because only 1% of variance in the outcome is explained, and thus, the theory's predictive ability is limited. As noted by Combs (2010),

> A theory might find support, but its explanatory power—that is, the effect size observed—is so weak that further efforts to develop the theory might not be warranted. . . . Small effects also raise questions about managerial relevance. . . . If managers begin to act on theories that are supported by small effects, they are not likely to notice positive results even when they occur. (p. 11)

In short, the magnitude of a meta-analytically derived effect size, expressed in absolute value, which denotes explanatory potential, is an important indicator of the practical usefulness of a theory (Aguinis, Werner, Abbott, Angert, Park, & Kohlhausen, 2010; Bacharach, 1989). Of course, even small effect sizes may have important practical consequences in some contexts (Cortina & Landis, 2009). However, regardless of their practical impact, theories that explain a larger portion of the variance in relevant outcomes are more useful than those that explain a small portion (Aguinis, Werner, et al., 2010; Bacharach, 1989).

Third, we examine the extent to which various methodological choices and judgment calls are associated with citation rates, which are considered an important indicator of scholarly influence and impact on the field (Podsakoff, MacKenzie, Podsakoff, & Bachrach, 2008). Moreover, we investigate whether effect sizes are related to their impact on the field (i.e., citations). An analysis of whether meta-analyses reporting larger effect sizes are cited more frequently compared with those reporting smaller effect sizes allows us to understand the

level of maturity of the field (Boyd, Finkelstein, & Gove, 2005; Long & Fox, 1995). Specifically, if citations are related to effect sizes, there would be evidence for scientific universalism: Published meta-analyses are cited according to their relative merit (i.e., ability to explain outcomes of interest). Instead, if citations are not related to effect sizes but to methodological choices and judgment calls not strongly related to the resulting effect sizes, there would be evidence for scientific particularism: Published meta-analyses are cited based on factors other than merit (e.g., extrascientific factors such as the popularity of certain choices and judgment calls).

Fourth, we investigate the extent to which meta-analyses have attempted to build new theories or test existing theories and whether these goals are related to the magnitude of meta-analytically derived effect sizes. We test the fit of a multivariate model that includes relationships among theory-building and theory-testing goals, magnitude of the meta-analytically derived effect sizes, year of publication of the meta-analysis, and citations. The test of this model enables us to understand empirically the interplay between methodological and theoretical issues, as well as scholarly impact (cf. Van Maanen, Sørensen, & Mitchell, 2007). Taken together, the investigation of these issues gives us a comprehensive data-based understanding of the choices that researchers make in conducting meta-analytic reviews and the implications of these practices for theory building and testing, the size of the obtained effect (i.e., an indicator of a theory's predictive potential and practical usefulness), and scholarly influence.

Method

Literature Search

As an initial step, we conducted a bibliographic search of the PsychINFO, ABI/Inform, and Academic Search (EBSCO) databases using keywords such as *meta-analysis, meta-analyses, meta-analytic, synthesis, syntheses,* and *quantitative review*. As an additional step to ensure the comprehensiveness of our review, we followed a procedure labeled *ancestry searching* (also known as citation chasing; e.g., Bays, 2001; Cooper, 1982). By working from the more contemporary references for meta-analysis, tracking these references for the prior meta-analytic work on which they relied, and iteratively continuing this process, it is possible to identify a set of common early references with no published predecessors. The ancestry approach can be quite productive. In our study, it provided a subset of studies the titles of which, and in some cases even the abstracts, would not have indicated unambiguously that the article relied on meta-analytic methodology. Examples of these include "Self-Efficacy and Work-Related Performance: The Integral Role of Individual Differences" (Judge, Jackson, Shaw, Scott, & Rich, 2007) and "Job and Life Satisfaction: A Reevaluation of the Strength of the Relationship and Gender Effects as a Function of the Date of the Study" (Tait, Padgett, & Baldwin, 1989).

As noted, our literature search covered the period from January 1982 through August 2009. While the starting date of any literature review is somewhat arbitrary (e.g., Cascio & Aguinis, 2008), our starting date was dictated by the year of the introduction of various meta-analytic approaches to the synthesis of data from separate studies (e.g., Hedges, 1982;

Hunter, Schmidt, & Jackson, 1982; Rosenthal & Rubin, 1982; see also Schmidt & Hunter, 2003, for a detailed historical account).

Criteria for Study Inclusion

To be considered for inclusion, an article had to meet several broad criteria. The article must have reported the fundamental data—number of studies, sample size of the studies, the effect size—on which the meta-analysis was based. Also, our review included meta-analyses addressing organizational topics only. Combining our literature search procedures with our study inclusion criteria resulted in 196 meta-analyses (24 in *AMJ*, 106 in *JAP*, 14 in *JOM,* 48 in *PPsych,* and 4 in *SMJ*), including a total of 5,581 effect sizes regarding the overall direct relationship (i.e., not subgroup based) between two variables. Of the total number of effect sizes, $k_{AMJ} = 400$, $k_{JAP} = 3,607$, $k_{JOM} = 195$, $k_{PPsych} = 1,364$, and $k_{SMJ} = 15$. These results suggest that meta-analysis is more popular in the organizational behavior (OB) and human resource management (HRM) literature than in the strategy literature.

Our focus is on direct (i.e., main) relationships between two variables and not on conditions under which a particular relationship is stronger or weaker (i.e., moderating effects) or intervening factors that explain a relationship between two variables (i.e., mediating effects). Our focus on direct relationships, as opposed to moderating or mediating effects, was guided by three factors. First, we were interested in the presumed effects of methodological choices and judgment calls on the meta-analytically derived effect size between X and Y and not on the boundary conditions for a theory (i.e., how the meta-analytically derived effect size between X and Y varies across groups or conditions, reasons why X affects Y). Second, there is a vast literature pointing to the fact that moderating effects are quite small at both the primary level (e.g., Aguinis, Beaty, Boik, & Pierce, 2005) and meta-analytic level of analysis (e.g., Schmidt, Pearlman, & Hunter, 1980). Third, a logical sequence in a systematic research agenda to understand the impact of meta-analytic methodological choices would be to focus on direct relationships first and to focus on other, more complex, effects at a later stage. This is because results of meta-analytic reviews are used for drawing conclusions about (a) direct relationships and (b) other types of relationships including moderating and mediating effects. All meta-analyses examine direct relationships, and the procedure for doing so is usually similar across studies (i.e., computation of a bivariate effect size estimate). However, not all meta-analyses examine moderating and mediating effects. Moreover, tests for possible mediation include a variety of approaches and are increasingly relying on combining meta-analytically derived results with other analytic approaches. For example, a meta-analytically derived correlation matrix can be used as input in a subsequent structural equation model (Colquitt et al., 2000). Thus, given these reasons, an examination of the impact of meta-analytic methodological choices on conclusions regarding moderating and mediating effects can be conducted in the future with the goal of gathering external validity evidence as an extension or constructive replication of the present study focusing on direct relationships.

Each meta-analysis reported multiple effect sizes. Within any given meta-analysis, some effect sizes are based on the same predictor or independent variable (e.g., job satisfaction)

but on a different criterion or dependent variable (e.g., task performance vs. contextual performance). In other cases, there may be overlap across effect sizes in terms of using the same criterion or dependent variable (e.g., supervisory ratings of performance) but a different predictor or independent variable (e.g., general mental ability vs. interview score). In yet many other cases, an article reports effect sizes computed on completely different predictors or independent variables and criteria or dependent variables. So, although in some cases there is overlap in terms of the variables used to compute some of the effect sizes within a given article, the majority of effect sizes do not share the measures and constructs underlying these measures. Thus, it would not be appropriate to aggregate effect sizes within articles. Accordingly, we conducted analyses at the effect-size level instead of the article level. Note that reviews of other types of effect sizes have followed the same approach (e.g., Aguinis et al., 2005).

Coding Process, Coding Fidelity, and Data Accuracy Checks

Variables coded for each of the meta-analyses included 21 methodological choices and judgment calls related to the various sequential steps involved in conducting a meta-analysis (cf. Bobko & Roth, 2008; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Wanous et al., 1989). Table 1 shows the 21 methodological choices and judgment calls. We included variables ranging from the stage of primary-level study retrieval (e.g., Were any studies eliminated and why?) to the data-analysis stage (i.e., Was a fixed- or a random-effects model used?), and finally, to the results reporting stage (i.e., Was a fail-safe test conducted?). We use the phrase *methodological choices and judgment calls* to refer to procedural decisions that meta-analysts make in conducting a meta-analysis. As in any content analysis, the choice for which variables to code is determined by (a) the need to address the questions of interest in any particular study and (b) practical constraints in terms of researchers' ability to extract the necessary data from the published articles (Duriau, Reger, & Pfarrer, 2007). Our approach was to be as inclusive as possible and extract as much information as possible from each article. However, we had to confront the fact that not all the information about methodological choices and judgment calls we could possibly code was reported by authors and, hence, available. Given documented difficulties in obtaining information from authors that is not available in the published articles (Aguinis et al., 2005), we coded as much information as was available in the articles. There may be additional variables related to methodological choices and judgment calls that could potentially be coded if the information was available.

Our coding included several additional descriptive characteristics. Examples of these variables are names of the authors, name of the journal in which the meta-analysis was published, year of publication, and the earliest and latest year of publication for a study included in the meta-analysis. In addition, our coding included data from third-party sources (e.g., Web of Science) regarding the total number of citations, as well as yearly citations, received by each article.

Our coding also involved classifying each of the 196 meta-analyses according to topic area. We coded for topic area because certain choices and judgment calls may be more meaningful in some areas compared with others. For example, a choice to correct for the effects of range restriction may be justified and even expected in the area of personnel selection in

Table 1
Percentage of Effect Sizes Computed as a Result of Various
Methodological Choices and Judgment Calls

1. Were any studies eliminated from the meta-analytic database? (n = 5,541)	
Yes	57.2
No	42.1
2. If any studies were eliminated from the meta-analytic database, what was the first elimination criterion? (n = 2,046)	
Independence issues	21.8
Sample was too small	9.5
Missing information in primary-level study	2.8
Outlier	2.6
3. Was the list of primary-level studies used in the meta-analysis included in the article? (n = 5,551)	
Yes	86.4
No	7.8
Available from the author(s) upon request	5.2
4. Was the list of primary-level study effect sizes used in the meta-analysis included in the article? (n = 5,576)	
No	92.3
Yes	7.0
Available from the author(s) upon request	0.6
5. Meta-analytic model used (n = 5,289)	
Random effects	87.5
Fixed effects	7.3
6. Meta-analytic procedure used (n = 5,064)	
Hunter and Schmidt (1990, 2004)	83.5
Hedges and Olkin (1985)	3.2
Raju, Burke, Normand, and Langlois (1991)	3.1
Rosenthal and Rubin (1982)	0.8
7. Software used (n = 565)	
SAS [proc means]	4.9
Hunter-Schmidt	3.6
Dstat	1.6
8. Was range restriction for the dependent variable reported? (n = 5,541)	
No	88.9
Yes, and correction specified	9.5
Yes, but no correction noted	0.9
9. Were multiple range restriction values noted for the dependent variables? (n = 3,375)	
No	49.9
Yes, and corrections specified	10.2
Yes, but no correction noted	0.4
10. Was range restriction for the independent variable reported? (n = 5,568)	
No	87.4
Yes, and correction specified	11.6
Yes, but no correction noted	0.7
11. Were multiple range restriction values noted for the independent variables? (n = 3,062)	
No	45.6
Yes	9.2
12. Was the reliability for the dependent variable reported? (n = 5,560)	
Yes, and correction specified	51.2
No	31.9
Yes, but no correction noted	16.6

(continued)

Table 1 (continued)

13. Were multiple reliability values noted for the dependent variables? ($n = 4,696$)	
No	39.5
Yes, and correction specified	34.2
Yes, but no correction noted	10.4
14. Was the reliability for the independent variable reported? ($n = 5,346$)	
Yes, and correction specified	48.3
No	31.9
Yes, but no correction noted	15.6
15. Were multiple reliability values noted for the independent variables? ($n = 4,636$)	
No	38.7
Yes, and correction specified	35.6
Yes, but no correction noted	8.8
16. Were obtained effect sizes used in follow-up analyses (e.g., multiple regression, structural equation modeling)? ($n = 5,581$)	
Yes	52.8
No	47.2
17. Type of effect size reported ($n = 5,577$)	
r	90.0
d	10.0
18. Was a confidence interval reported? ($n = 5,539$)	
Yes, 95	47.0
No	43.8
Yes, 90	7.3
19. Was a credibility interval reported? ($n = 5,455$)	
No	63.6
Yes, 90	12.8
Yes, 80	10.7
Yes, 95	10.6
20. Was a fail-safe statistic reported? ($n = 5,569$)	
No	92.7
Yes	7.1
21. Was there any other test of possible publication bias? ($n = 5,409$)	
No	94.9
Yes	2.0

Note: $N = 5,581$ meta-analytically derived effect-size estimates; $n =$ number of effect size estimates (out of the total of 5,581) for which information was available to code for each of the 21 choices and judgment calls. Percentages do not total 100 in all cases, due to missing data.

which validity coefficients between preemployment tests and measures of job performance are typically affected due to selection artifacts (Aguinis, Culpepper, & Pierce, in press). However, the choice to implement range restriction corrections may not lead to substantive differences in the resulting effect sizes in other areas of research in which selection artifacts are not as pervasive. As the first step in our classification, we used the 15-category taxonomy for OB/HRM topics from Cascio and Aguinis (2008) and the 8-category taxonomy for strategy based on the Strategic Management Society's interest groups (http://strategicmanage ment.net/ig/index.php). Some of the categories had zero meta-analyses, so we merged them with others. Moreover, given that the majority of meta-analyses addressed OB/HRM topics, our final 12-category taxonomy consists of the following: (1) job analysis ($k = 9$); (2)

research methodology and psychometric issues ($k = 27$); (3) predictors of performance ($k = 32$); (4) performance measurement and work outcomes ($k = 24$); (5) training and development ($k = 11$); (6) industrial relations, reward systems, career issues, and decision making ($k = 10$); (7) work motivation and attitudes ($k = 34$); (8) leader influences ($k = 11$); (9) work groups and teams ($k = 14$); (10) societal issues ($k = 7$); (11) corporate strategy and governance ($k = 7$); and (12) all other macro/strategy topics ($k = 10$).

Finally, our coding also involved ratings of the extent to which each effect size attempted to build a new theory or test existing theory. Colquitt and Zapata-Phelan (2007) conducted a content analysis of the 667 empirical articles published in *AMJ* from 1963 to 2007 to understand whether these studies were mainly attempting to build new theory or test an existing theory. Colquitt and Zapata-Phelan defined theory building as "the degree to which an empirical article clarifies or supplements existing theory or introduces relationships and constructs that serve as the foundations for new theory" (p. 1283). Alternatively, they defined theory testing as "the degree to which existing theory is applied in an empirical study as a means of grounding a specific set of a priori hypotheses" (p. 1284). Following their data-coding protocol, we coded each of the 5,581 effect sizes in our database along the theory-building and theory-testing dimensions using the following scales offered by Colquitt and Zapata-Phelan: (1) Building new theory: 1 = *attempts to replicate previously demonstrated effects,* 2 = *examines effects that have been the subject of prior theorizing,* 3 = *introduces a new mediator or moderator of an existing relationship or process,* 4 = *examines a previously unexplored relationship or process,* or 5 = *introduces a new construct* (or significantly reconceptualizes an existing one). (2) Testing existing theory: 1 = *is inductive or grounds predictions with logical speculation,* 2 = *grounds predictions with references to past findings,* 3 = *grounds predictions with existing conceptual arguments,* 4 = *grounds predictions with existing models, diagrams, or figures,* or 5 = *grounds predictions with existing theory*.

We created a coding protocol to guide the process of capturing each variable value (cf. Duriau et al., 2007). The protocol included the definition of each variable and general guidelines regarding the assignment of values. For objective variables (e.g., year of publication, number of studies included in a meta-analysis), we conducted analyses based on descriptive statistics and graphic displays (e.g., minimum and maximum values, variances, frequency distributions) to identify both out-of-range and unlikely values. In all cases, the articles were reviewed and such errors in the database were subsequently corrected. For subjective variables (e.g., the extent to which the goal was to build new theory or test an existing theory), our accuracy-check process involved additional steps. First, we created a coding sheet with Microsoft Excel functions, including mouse-over variable cell comments and mouse-click drop-down menus. Second, coding was conducted independently by two members of the authorship team in geographically dispersed locations. Third, we randomly selected subsets of 10 meta-analyses to conduct agreement checks. The objective was to determine where there were inconsistencies in the data independently entered by one of the coders when compared with data entered by the other coder for the same cases. In the first of these evaluations, the consistency rate was 95.4%. The second resulted in a consistency rate of 96.8%. Through these iterations, we were able to determine the source of most of the inconsistencies. For the final quality check, agreement was 99.7%. Throughout the process, in the few cases for which there was no consensus, disagreements were resolved

Figure 1
Number of Meta-Analytically Derived Effect Sizes Between
January 1982 and December 2008 and Quadratic Trend for Frequency
of Effect Sizes Expected to Be Reported Until the Year 2020
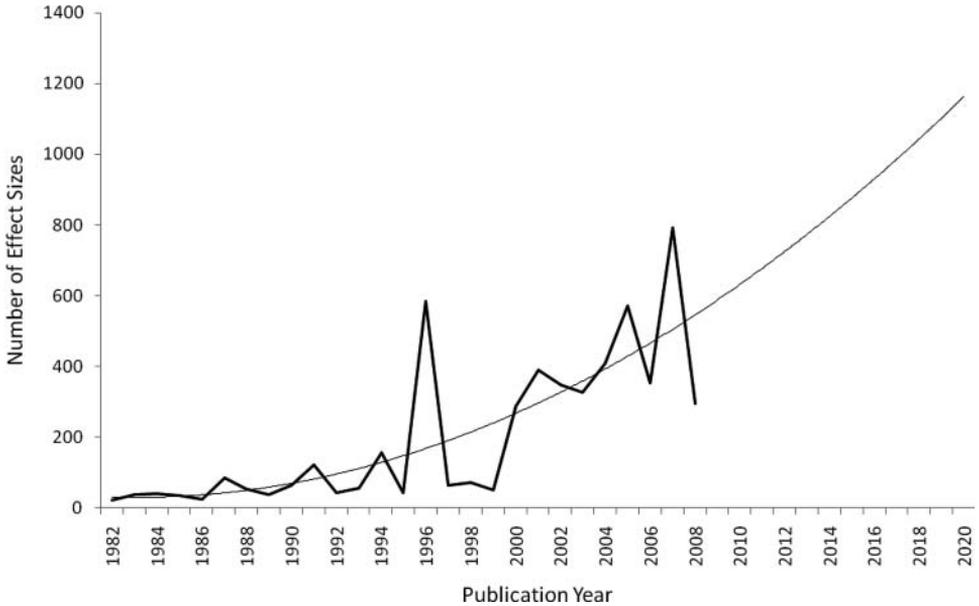
Note: Effect sizes reported in *Academy of Management Journal, Journal of Applied Psychology, Journal of Management, Personnel Psychology,* and *Strategic Management Journal*.

by revisiting the meta-analysis. In all cases, a consensual coding decision was reached prior to analyses.

The completion of our data-gathering process resulted in a rectangular matrix (i.e., 118 variables and 5,581 rows of values) with a total of 658,558 elements. The citations for the 196 meta-analyses included in our review are available online at http://mypage.iu.edu/~haguinis.

Results

Frequency of Meta-Analytically Derived Effect Sizes Over Time

Figure 1 shows the number of meta-analytically derived effect sizes reported in *AMJ, JAP, JOM, PPsych,* and *SMJ* from January 1982 through December 2008 (we did not include 2009 articles in this graph because our analysis included only part of 2009). As can be seen in this figure, the number of meta-analyses has increased steadily, particularly in the 21st century. The number of effect sizes reported annually since the year 2000 has been at least

286 (in 2000) and reached a high of 791 in 2007. In fact, there is a polynomial (i.e., qua-dratic) trend suggesting that the number of published meta-analyses is likely to continue to increase over the next decade. This trend, also shown in Figure 1, suggests that the expected number of meta-analytically derived effect sizes to be published annually will approach 1,000 around the year 2015 and 1,200 around the year 2020.

Typical Methodological Choices and Judgment Calls and Descriptive Statistics

Table 1 includes information regarding the most popular decisions meta-analysts make in terms of methodological choices and judgment calls. The following is a summary of the most typical characteristics of the published meta-analyses. First, in terms of design, studies are eliminated from the meta-analytic database based primarily on independence issues, and the list of studies used in the meta-analysis is included in the article, but the list of effect sizes used in the meta-analytic computations is not. Second, in terms of analysis, there is a preference for a random-effects model and the Hunter and Schmidt (1990, 2004) procedures. Overall, however, range restriction is not reported or corrected for in the independent or dependent variables. On the other hand, reliability is reported and measurement error is cor-rected for in the independent and dependent variables. Also, about half of the effect sizes are used in follow-up analyses (e.g., multiple regression, structural equation modeling). Third, in terms of reporting results, the preferred effect-size metric is *r,* which is accompanied with a 95% confidence interval in about 50% of cases but without a credibility interval. Finally, regarding robustness and replicability of results, the typical meta-analysis does not report fail-safe statistics or any other test of potential publication bias.

In addition to the information included in Table 1, across the 5,581 effect sizes, the mean absolute *r* value (i.e., we converted *d*s to *r*s when an article reported *d*s) is $|\bar{r}| = .26$ ($SD = .18$) and the median $|\bar{r}|$ is .23, suggesting positive skew (i.e., skewness = .85). Note that these effect-size statistics refer to corrected (for statistical and methodological artifacts) effects when the authors reported corrected effects, and they are uncorrected if authors reported uncorrected effects. The frequency distribution of the 5,581 effect sizes is included in Figure 2 (top panel).

Additional normative results are that each meta-analytically derived effect size was com-puted using a mean of 18 primary-level effects ($Mdn = 9$) and a mean total sample size of 12,885 ($Mdn = 2,244$). Also, when a measurement error correction was applied, the mean correction factor (i.e., reliability estimate) for the independent variable or predictor was .78 ($SD = .09$, $Mdn = .77$) and the mean correction factor for the dependent variable or criterion was .79 ($SD = .08$, $Mdn = .80$). These results support calls for better measurement instru-ments (Aguinis, Pierce, Bosco, & Muslin, 2009) given that about 20% of variance in both the independent and dependent variables is random error. In terms of scholarly impact, according to Web of Science, the 196 articles received a mean of 88.44 total citations ($Mdn = 45.00$) and a mean of 8.47 citations per year ($Mdn = 6.00$). Note that our analyses involving citations use citations per year and not total number of citations because older articles have a greater opportunity to be cited.

Figure 2
Frequency Distributions of Meta-Analytically Derived Effect Sizes
(top panel) and Citations Per Year (bottom panel)
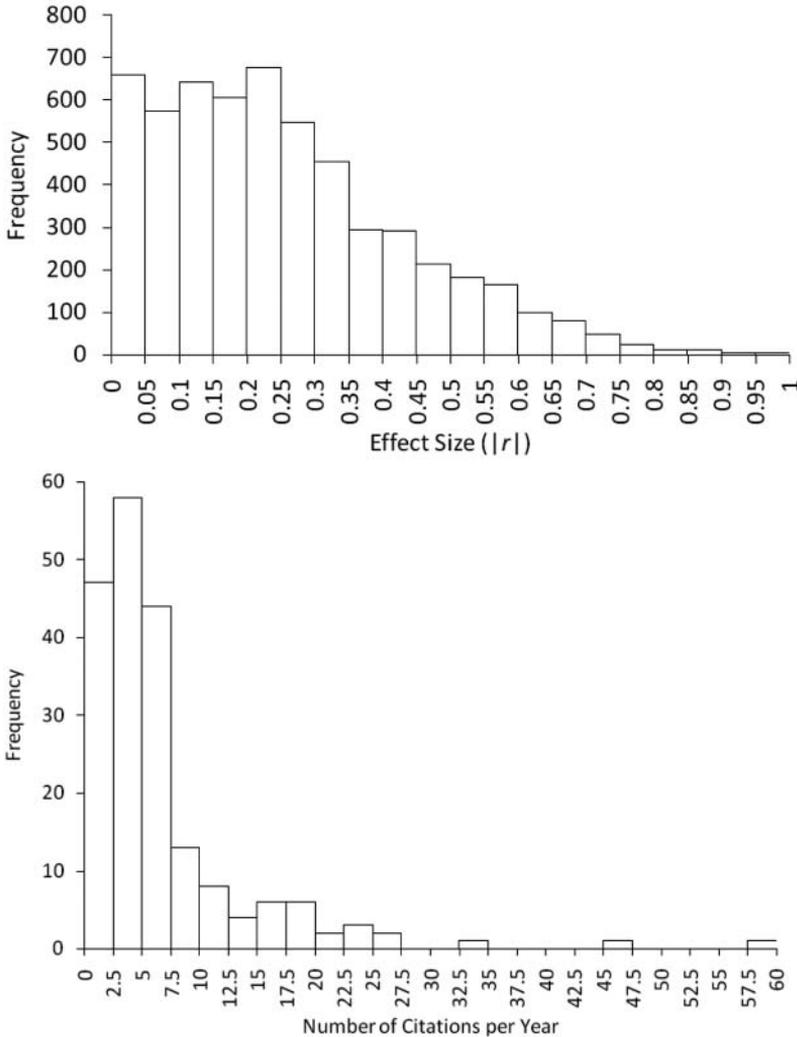
Finally, meta-analyses included primary-level studies published as early as 1917. The frequency distribution of citations per year received by the 196 meta-analyses is included in Figure 2 (bottom panel). Similar to other research focusing on citation analysis (Podsakoff et al., 2008), this figure shows that the distribution is positively skewed.

Relationship Between Methodological Choices and Judgment Calls With Effect Sizes

The goal of the next step in our analysis was to understand the relationship between each of the 21 methodological choices and judgment calls included in Table 1 and the resulting meta-analytically derived effect sizes. The main question addressed by this set of analyses is the extent to which the various choices and judgment calls actually have a putative effect on the size of the resulting effects.

To investigate the relationship between methodological choices and judgment calls and the obtained effect sizes, we conducted analyses of covariance (ANCOVAs) using methodological choices and judgment calls as independent variables and absolute-value meta-analytically derived effect sizes (i.e., $|r|$) as the dependent variable. The covariates were 11 dummy-coded variables representing the 12-category topic area taxonomy described in the Method section. In the analyses described below, the dependent variable is the corrected effect size when it was reported in the meta-analyses ($k = 4,852$) and the uncorrected effect size ($k = 729$) if corrected effect sizes were not reported.

Table 2 shows results of analyses relating methodological choices and judgment calls with the magnitude of the meta-analytically derived effects. Due to a total of 5,581 effect-size estimates and the resulting high level of statistical power, it is expected that most, if not all, of the comparisons would be statistically significant. Thus, given the large sample sizes, Table 2 includes information regarding the eta-squared value (η^2) for each comparison. Each η^2 value indicates the proportion of variance in the meta-analytically derived $|r|$s explained by each choice and judgment call (Pierce, Block, & Aguinis, 2004). The mean η^2 for the values shown in Table 2 is .007 and the median is .004. The largest η^2 is .071 (i.e., type of software used), the second largest value is .011 (i.e., whether range restriction was reported for the independent variable), and the third largest value is .009 (i.e., type of effect size reported). None of the remaining 18 choices and judgment calls accounted for more than 0.8% percent of variance in the meta-analytically derived effect sizes. In fact, the total proportion of variance explained by these 18 methodological choices and judgments calls combined is only .065. In other words, assuming these 18 methodological choices and judgment calls are fully independent, which is quite unlikely, they account for a maximum of only about 6.5% of the total variance in the meta-analytically derived effect sizes. So, the 6.5% value is probably overestimating the already small effect. In short, results indicate that the methodological choices and judgment calls we reviewed have little impact on the resulting magnitude of the meta-analytically derived effect sizes.

We also examined the impact of implementing methodological choices and judgment calls specifically addressing corrections for statistical and methodological artifacts such as measurement error and range restriction in the predictor and/or the criterion variables. Table 3 includes summary data from each of the 138 meta-analyses that reported both uncorrected and corrected effect sizes. As expected, corrected coefficients are larger than their uncorrected counterparts. Overall, the mean uncorrected effect size is .206 (*Mdn* = .190) and the mean corrected effect size is .261 (*Mdn* = .240). So, the mean corrected effect size is .055 (*Mdn* = .050) correlation coefficient units larger than its uncorrected counterpart. As shown in Table 3, differences between corrected and uncorrected summary effect sizes across the 12

Table 2
Relationship Between Methodological Choices and Judgment Calls With Absolute
Value of Meta-Analytically Derived Effect Sizes ($|r|$)

1. Elimination of studies from the meta-analytic database
   $F(1, 5539) = 0.44$, $p = .506$, $\eta^2 = .000$
   $|\bar{r}|_{\text{yes}} = .26$ ($Mdn = .24$, $SD = .17$, $n = 3,192$)
   $|\bar{r}|_{\text{no}} = .25$ ($Mdn = .21$, $SD = .19$, $n = 2,349$)

2. If any studies were eliminated from the meta-analytic database, elimination criterion
   $F(3, 2042) = 3.66$, $p = .012$, $\eta^2 = .005$
   $|\bar{r}|_{\text{independence issues}} = .28$ ($Mdn = .27$, $SD = .18$, $n = 1,216$)
   $|\bar{r}|_{\text{sample was too small}} = .27$ ($Mdn = .25$, $SD = .17$, $n = 530$)
   $|\bar{r}|_{\text{missing information}} = .25$ ($Mdn = .22$, $SD = .15$, $n = 157$)
   $|\bar{r}|_{\text{outliers}} = .27$ ($Mdn = .21$, $SD = .17$, $n = 143$)

3. Inclusion of the list of primary-level studies used in the meta-analysis included in the article
   $F(2, 5548) = 19.81$, $p = .000$, $\eta^2 = .007$
   $|\bar{r}|_{\text{yes}} = .26$ ($Mdn = .22$, $SD = .18$, $n = 4,821$)
   $|\bar{r}|_{\text{no}} = .22$ ($Mdn = .21$, $SD = .15$, $n = 435$)
   $|\bar{r}|_{\text{available from the author(s) upon request}} = .30$ ($Mdn = .29$, $SD = .17$, $n = 295$)

4. Was the list of primary-level study effect sizes used in the meta-analysis included in the article?
   $F(2, 5573) = 12.45$, $p = .000$, $\eta^2 = .004$
   $|\bar{r}|_{\text{no}} = .25$ ($Mdn = .22$, $SD = .18$, $n = 5,151$)
   $|\bar{r}|_{\text{yes}} = .30$ ($Mdn = .28$, $SD = .16$, $n = 390$)
   $|\bar{r}|_{\text{available from the author(s) upon request}} = .29$ ($Mdn = .14$, $SD = .31$, $n = 35$)

5. Meta-analytic model used
   $F(1, 5287) = 0.64$, $p = .422$, $\eta^2 = .000$
   $|\bar{r}|_{\text{random effects}} = .26$ ($Mdn = .23$, $SD = .18$, $n = 4,882$)
   $|\bar{r}|_{\text{fixed effects}} = .27$ ($Mdn = .22$, $SD = .19$, $n = 407$)

6. Meta-analytic procedure used
   $F(3, 5060) = 5.81$, $p = .001$, $\eta^2 = .003$
   $|\bar{r}|_{\text{Hunter and Schmidt (1990, 2004)}} = .26$ ($Mdn = .23$, $SD = .18$, $n = 4,661$)
   $|\bar{r}|_{\text{Hedges and Olkin (1985)}} = .24$ ($Mdn = .17$, $SD = .22$, $n = 181$)
   $|\bar{r}|_{\text{Rosenthal and Rubin (1982)}} = .33$ ($Mdn = .31$, $SD = .15$, $n = 47$)
   $|\bar{r}|_{\text{Raju, Burke, Normand, and Langlois (1991)}} = .29$ ($Mdn = .27$, $SD = .16$, $n = 175$)

7. Software used
   $F(2, 562) = 21.17$, $p = .000$, $\eta^2 = .071$
   $|\bar{r}|_{\text{SAS [proc means]}} = .30$ ($Mdn = .27$, $SD = .16$, $n = 272$)
   $|\bar{r}|_{\text{Hunter-Schmidt}} = .18$ ($Mdn = .13$, $SD = .14$, $n = 202$)
   $|\bar{r}|_{\text{Dstat}} = .35$ ($Mdn = .30$, $SD = .22$, $n = 91$)

8. Was range restriction for the dependent variable reported?
   $F(2, 5538) = 13.91$, $p = .000$, $\eta^2 = .005$
   $|\bar{r}|_{\text{no}} = .25$ ($Mdn = .22$, $SD = .18$, $n = 4,961$)
   $|\bar{r}|_{\text{yes, and correction specified}} = .28$ ($Mdn = .25$, $SD = .20$, $n = 529$)
   $|\bar{r}|_{\text{yes, but no correction noted}} = .30$ ($Mdn = .32$, $SD = .12$, $n = 51$)

9. Were multiple range restriction values noted for the dependent variables?
   $F(2, 3372) = 14.11$, $p = .000$, $\eta^2 = .008$
   $|\bar{r}|_{\text{no}} = .25$ ($Mdn = .22$, $SD = .18$, $n = 2,785$)
   $|\bar{r}|_{\text{yes, and correction specified}} = .28$ ($Mdn = .25$, $SD = .20$, $n = 570$)
   $|\bar{r}|_{\text{yes, but no correction noted}} = .28$ ($Mdn = .32$, $SD = .10$, $n = 20$)

10. Was range restriction for the independent variable reported?
   $F(2, 5565) = 32.04$, $p = .000$, $\eta^2 = .011$
   $|\bar{r}|_{\text{no}} = .25$ ($Mdn = .22$, $SD = .18$, $n = 4,879$)
   $|\bar{r}|_{\text{yes, and correction specified}} = .26$ ($Mdn = .27$, $SD = .19$, $n = 650$)
   $|\bar{r}|_{\text{yes, but no correction noted}} = .35$ ($Mdn = .35$, $SD = .17$, $n = 39$)

(continued)

Table 2 (continued)

11. Were multiple range restriction values noted for the independent variables?
    $F(1, 3060) = 14.89$, $p = .000$, $\eta^2 = .005$
    $|\bar{r}|_{no} = .25$ ($Mdn = .22$, $SD = .17$, $n = 2,547$)
    $|\bar{r}|_{yes, and correction specified} = .28$ ($Mdn = .25$, $SD = .20$, $n = 515$)
12. Was the reliability for the dependent variable reported?
    $F(2, 5557) = 11.85$, $p = .000$, $\eta^2 = .004$
    $|\bar{r}|_{yes, and correction specified} = .26$ ($Mdn = .22$, $SD = .18$, $n = 2,856$)
    $|\bar{r}|_{no} = .25$ ($Mdn = .22$, $SD = .17$, $n = 1,780$)
    $|\bar{r}|_{yes, but no correction noted} = .27$ ($Mdn = .25$, $SD = .18$, $n = 924$)
13. Were multiple reliability values noted for the dependent variables?
    $F(2, 4693) = 6.11$, $p = .002$, $\eta^2 = .003$
    $|\bar{r}|_{no} = .26$ ($Mdn = .23$, $SD = .18$, $n = 2,207$)
    $|\bar{r}|_{yes, and correction specified} = .26$ ($Mdn = .23$, $SD = .19$, $n = 1,910$)
    $|\bar{r}|_{yes, but no correction noted} = .26$ ($Mdn = .24$, $SD = .17$, $n = 579$)
14. Was the reliability for the independent variable reported?
    $F(2, 5343) = 14.30$, $p = .000$, $\eta^2 = .005$
    $|\bar{r}|_{yes, and correction specified} = .26$ ($Mdn = .23$, $SD = .18$, $n = 2,693$)
    $|\bar{r}|_{no} = .25$ ($Mdn = .22$, $SD = .17$, $n = 1,781$)
    $|\bar{r}|_{yes, but no correction noted} = .28$ ($Mdn = .25$, $SD = .18$, $n = 872$)
15. Were multiple reliability values noted for the independent variables?
    $F(2, 4633) = 10.19$, $p = .000$, $\eta^2 = .004$
    $|\bar{r}|_{no} = .26$ ($Mdn = .23$, $SD = .17$, $n = 2,159$)
    $|\bar{r}|_{yes, and correction specified} = .26$ ($Mdn = .23$, $SD = .19$, $n = 1,985$)
    $|\bar{r}|_{yes, but no correction noted} = .26$ ($Mdn = .24$, $SD = .18$, $n = 492$)
16. Were obtained effect sizes used in follow-up analyses (e.g., multiple regression, structural equation modeling)?
    $F(1, 5579) = 0.73$, $p = .392$, $\eta^2 = .000$
    $|\bar{r}|_{yes} = .26$ ($Mdn = .23$, $SD = .18$, $n = 2,944$)
    $|\bar{r}|_{no} = .25$ ($Mdn = .22$, $SD = .18$, $n = 2,637$)
17. Type of effect size reported
    $F(1, 5575) = 51.96$, $p = .000$, $\eta^2 = .009$
    $|\bar{r}|_{r} = .26$ ($Mdn = .23$, $SD = .18$, $n = 5,021$)
    $|\bar{r}|_{d} = .22$ ($Mdn = .21$, $SD = .15$, $n = 556$)
18. Was a confidence interval reported?
    $F(2, 5468) = 21.41$, $p = .000$, $\eta^2 = .008$
    $|\bar{r}|_{yes, 95\%} = .27$ ($Mdn = .24$, $SD = .18$, $n = 2,621$)
    $|\bar{r}|_{no} = .24$ ($Mdn = .22$, $SD = .18$, $n = 2,444$)
    $|\bar{r}|_{yes, 90\%} = .22$ ($Mdn = .19$, $SD = .15$, $n = 406$)
19. Was a credibility interval reported?
    $F(3, 5451) = 5.99$, $p = .000$, $\eta^2 = .003$
    $|\bar{r}|_{no} = .26$ ($Mdn = .23$, $SD = .18$, $n = 3,548$)
    $|\bar{r}|_{yes, 90\%} = .24$ ($Mdn = .21$, $SD = .18$, $n = 714$)
    $|\bar{r}|_{yes, 95\%} = .26$ ($Mdn = .23$, $SD = .17$, $n = 594$)
    $|\bar{r}|_{yes, 80\%} = .28$ ($Mdn = .25$, $SD = .18$, $n = 599$)
20. Was a fail-safe statistic reported?
    $F(1, 5567) = 0.34$, $p = .559$, $\eta^2 = .000$
    $|\bar{r}|_{no} = .26$ ($Mdn = .23$, $SD = .18$, $n = 5,173$)
    $|\bar{r}|_{yes} = .26$ ($Mdn = .22$, $SD = .16$, $n = 396$)
21. Was there any other test of possible publication bias?
    $F(1, 5407) = 5.86$, $p = .016$, $\eta^2 = .001$
    $|\bar{r}|_{no} = .26$ ($Mdn = .23$, $SD = .18$, $n = 5,296$)
    $|\bar{r}|_{yes} = .19$ ($Mdn = .17$, $SD = .12$, $n = 113$)

Note: $|\bar{r}|$ = mean absolute meta-analytically derived effect size. $N = 5,581$ meta-analytically derived effect-size estimates; n = number of effect size estimates (out of the total of 5,581) for which information was available to code for each of the 21 choices and judgment calls.

Table 3
Impact of Implementing Corrections for Statistical and Methodological
Artifacts on the Resulting Effect Sizes Based on Meta-Analyses
Reporting Both Uncorrected and Corrected Effect Sizes

Topic Area	N (k)	$|\bar{r}|_u$	$|\bar{r}|_c$	$|\bar{r}|_c - |\bar{r}|_u$	$|\bar{r}|_{u(Mdn)}$	$|\bar{r}|_{c(Mdn)}$	$|\bar{r}|_{c(Mdn)} - |\bar{r}|_{u(Mdn)}$
(All topic areas)	3,912 (138)	.206	.261	.055	.190	.240	.050
Job analysis	566 (7)	.221	.291	.071	.210	.260	.050
Research methodology and psychometric issues	305 (16)	.206	.299	.094	.200	.290	.090
Predictors of performance	729 (25)	.195	.250	.055	.180	.240	.060
Performance measurement and work outcomes	219 (17)	.213	.268	.055	.170	.240	.070
Training and development	243 (8)	.204	.257	.054	.180	.230	.050
Industrial relations, reward systems, career issues, and decision making	242 (9)	.254	.310	.056	.230	.290	.060
Work motivation and attitudes	662 (25)	.222	.266	.043	.200	.230	.030
Leader influences	578 (11)	.172	.211	.039	.130	.160	.030
Work groups and teams	115 (10)	.225	.263	.038	.220	.260	.040
Societal issues	221 (5)	.199	.241	.042	.190	.230	.040
All other macro/strategy topics	18 (1)	.106	.134	.028	.085	.115	.030
Corporate strategy and governance	14 (4)	.061	.089	.028	.040	.075	.035

Note: $|\bar{r}|_u$ = uncorrected mean absolute effect size; $|\bar{r}|_c$ = corrected mean absolute effect size; $|\bar{r}|_{u(Mdn)}$ = uncorrected median absolute effect size; $|\bar{r}|_{c(Mdn)}$ = corrected median absolute effect size. N = number of meta-analytically derived effect-size estimates for which both uncorrected and corrected values were reported; k = number of articles reporting both corrected and uncorrected effect sizes.

topic areas range from about .03 to .09 correlation coefficient units. Stated differently, using corrections improves the proportion of variance explained in outcomes by an average of about .25%, and this improvement ranges from about .09% to .81% across the 12 topic areas.

Relationship Between Methodological Choices and Judgment Calls With Citations Per Year

As a next step in our data-analytic strategy, we report results to address the question of whether adopting certain methodological choices and judgment calls is related to the impact of a meta-analysis on the field, as indicated by the number of citations received per year since the meta-analysis was published. Table 4 includes results of ANCOVAs using each of the 21 methodological choices and judgment calls as independent variables and the number

Table 4
Relationship Between Methodological Choices and
Judgment Calls With Citations per Year

1. Elimination of studies from the meta-analytic database
    $F(1, 5539) = 9.55$, $p = .002$, $\eta^2 = .002$
    $\overline{C}_{yes} = 9.67$ ($Mdn = 6.57$, $SD = 9.27$, $n = 3,192$)
    $\overline{C}_{no} = 6.94$ ($Mdn = 4.14$, $SD = 10.69$, $n = 2,349$)
2. If any studies were eliminated from the meta-analytic database, elimination criterion
    $F(3, 2042) = 71.47$, $p = .000$, $\eta^2 = .095$
    $\overline{C}_{independence\ issues} = 11.60$ ($Mdn = 6.57$, $SD = 11.66$, $n = 1,216$)
    $\overline{C}_{sample\ was\ too\ small} = 9.96$ ($Mdn = 7.67$, $SD = 5.49$, $n = 530$)
    $\overline{C}_{missing\ information} = 18.11$ ($Mdn = 24.71$, $SD = 9.17$, $n = 157$)
    $\overline{C}_{outliers} = 6.20$ ($Mdn = 4.00$, $SD = 5.78$, $n = 143$)
3. Inclusion of the list of primary-level studies used in the meta-analysis included in the article
    $F(2, 5548) = 125.19$, $p = .000$, $\eta^2 = .043$
    $\overline{C}_{yes} = 7.83$ ($Mdn = 5.14$, $SD = 8.19$, $n = 4,821$)
    $\overline{C}_{no} = 15.27$ ($Mdn = 4.71$, $SD = 20.84$, $n = 435$)
    $\overline{C}_{available\ from\ the\ author(s)\ upon\ request} = 9.55$ ($Mdn = 6.57$, $SD = 8.05$, $n = 295$)
4. Was the list of primary-level study effect sizes used in the meta-analysis included in the article?
    $F(2, 5573) = 3.21$, $p = .040$, $\eta^2 = .001$
    $\overline{C}_{no} = 8.45$ ($Mdn = 6.00$, $SD = 10.02$, $n = 5,151$)
    $\overline{C}_{yes} = 8.89$ ($Mdn = 3.92$, $SD = 9.80$, $n = 390$)
    $\overline{C}_{available\ from\ the\ author(s)\ upon\ request} = 6.37$ ($Mdn = 6.08$, $SD = 0.91$, $n = 35$)
5. Meta-analytic model used
    $F(1, 5287) = 161.90$, $p = .000$, $\eta^2 = .030$
    $\overline{C}_{random\ effects} = 9.01$ ($Mdn = 6.20$, $SD = 10.45$, $n = 4,882$)
    $\overline{C}_{fixed\ effects} = 3.87$ ($Mdn = 2.67$, $SD = 3.99$, $n = 407$)
6. Meta-analytic procedure used
    $F(3, 5060) = 52.49$, $p = .000$, $\eta^2 = .030$
    $\overline{C}_{Hunter\ and\ Schmidt\ (1990,\ 2004)} = 9.03$ ($Mdn = 6.20$, $SD = 10.41$, $n = 4,661$)
    $\overline{C}_{Hedges\ and\ Olkin\ (1985)} = 5.54$ ($Mdn = 5.00$, $SD = 5.24$, $n = 181$)
    $\overline{C}_{Rosenthal\ and\ Rubin\ (1982)} = 4.93$ ($Mdn = 4.93$, $SD = 0.00$, $n = 47$)
    $\overline{C}_{Raju,\ Burke,\ Normand,\ and\ Langlois\ (1991)} = .94$ ($Mdn = 0.00$, $SD = 10.17$, $n = 175$)
7. Software used
    $F(2, 562) = 111.81$, $p = .000$, $\eta^2 = .287$
    $\overline{C}_{SAS\ [proc\ means]} = 4.90$ ($Mdn = 5.00$, $SD = 1.43$, $n = 272$)
    $\overline{C}_{Hunter-Schmidt} = 27.92$ ($Mdn = 9.33$, $SD = 24.83$, $n = 202$)
    $\overline{C}_{Dstat} = 20.59$ ($Mdn = 27.00$, $SD = 8.11$, $n = 91$)
8. Was range restriction for the dependent variable reported?
    $F(2, 5538) = 194.88$, $p = .000$, $\eta^2 = .066$
    $\overline{C}_{no} = 7.77$ ($Mdn = 5.00$, $SD = 8.34$, $n = 4,961$)
    $\overline{C}_{yes,\ and\ correction\ specified} = 15.89$ ($Mdn = 7.67$, $SD = 18.20$, $n = 529$)
    $\overline{C}_{yes,\ but\ no\ correction\ noted} = 3.19$ ($Mdn = 2.85$, $SD = 2.17$, $n = 51$)
9. Were multiple range restriction values noted for the dependent variables?
    $F(2, 3372) = 85.68$, $p = .000$, $\eta^2 = .049$
    $\overline{C}_{no} = 8.11$ ($Mdn = 6.11$, $SD = 9.26$, $n = 2,785$)
    $\overline{C}_{yes,\ and\ correction\ specified} = 14.92$ ($Mdn = 7.67$, $SD = 17.89$, $n = 570$)
    $\overline{C}_{yes,\ but\ no\ correction\ noted} = 2.28$ ($Mdn = 2.85$, $SD = 1.17$, $n = 20$)
10. Was range restriction for the independent variable reported?
    $F(2, 5565) = 124.52$, $p = .000$, $\eta^2 = .043$
    $\overline{C}_{no} = 7.77$ ($Mdn = 5.00$, $SD = 8.38$, $n = 4,879$)
    $\overline{C}_{yes,\ and\ correction\ specified} = 14.11$ ($Mdn = 7.67$, $SD = 16.96$, $n = 650$)
    $\overline{C}_{yes,\ but\ no\ correction\ noted} = 1.67$ ($Mdn = 1.86$, $SD = 0.57$, $n = 39$)

(continued)

Table 4 (continued)

---

11. Were multiple range restriction values noted for the independent variables?

$F(1, 3060) = 221.65$, $p = .000$, $\eta^2 = .068$

$\overline{C}_{\text{no}} = 8.85$ ($Mdn = 6.20$, $SD = 9.43$, $n = 2,547$)

$\overline{C}_{\text{yes, and correction specified}} = 16.27$ ($Mdn = 7.67$, $SD = 18.29$, $n = 515$)

12. Was the reliability for the dependent variable reported?

$F(2, 5557) = .80$, $p = .448$, $\eta^2 = .000$

$\overline{C}_{\text{yes, and correction specified}} = 8.61$ ($Mdn = 6.20$, $SD = 9.83$, $n = 2,856$)

$\overline{C}_{\text{no}} = 8.83$ ($Mdn = 6.11$, $SD = 11.10$, $n = 1,780$)

$\overline{C}_{\text{yes, but no correction noted}} = 7.46$ ($Mdn = 3.29$, $SD = 7.87$, $n = 924$)

13. Were multiple reliability values noted for the dependent variables?

$F(2, 4693) = 28.52$, $p = .000$, $\eta^2 = .012$

$\overline{C}_{\text{no}} = 9.57$ ($Mdn = 6.11$, $SD = 6.23$, $n = 2,207$)

$\overline{C}_{\text{yes, and correction specified}} = 10.02$ ($Mdn = 7.67$, $SD = 11.22$, $n = 1,910$)

$\overline{C}_{\text{yes, but no correction noted}} = 4.95$ ($Mdn = 2.33$, $SD = 6.23$, $n = 579$)

14. Was the reliability for the independent variable reported?

$F(2, 5343) = 3.92$, $p = .020$, $\eta^2 = .001$

$\overline{C}_{\text{yes, and correction specified}} = 8.01$ ($Mdn = 6.20$, $SD = 9.63$, $n = 2,693$)

$\overline{C}_{\text{no}} = 8.70$ ($Mdn = 5.71$, $SD = 11.03$, $n = 1,781$)

$\overline{C}_{\text{yes, but no correction noted}} = 7.77$ ($Mdn = 4.00$, $SD = 8.19$, $n = 872$)

15. Were multiple reliability values noted for the independent variables?

$F(2, 4633) = 39.69$, $p = .000$, $\eta^2 = .017$

$\overline{C}_{\text{no}} = 9.78$ ($Mdn = 6.11$, $SD = 10.66$, $n = 2,159$)

$\overline{C}_{\text{yes, and correction specified}} = 9.61$ ($Mdn = 7.56$, $SD = 11.16$, $n = 1,985$)

$\overline{C}_{\text{yes, but no correction noted}} = 5.73$ ($Mdn = 2.33$, $SD = 6.93$, $n = 492$)

16. Were obtained effect sizes used in follow-up analyses (e.g., multiple regression, structural equation modeling)?

$F(1, 5579) = 17.71$, $p = .000$, $\eta^2 = .003$

$\overline{C}_{\text{yes}} = 7.72$ ($Mdn = 4.63$, $SD = 9.39$, $n = 2,944$)

$\overline{C}_{\text{no}} = 9.29$ ($Mdn = 7.00$, $SD = 10.51$, $n = 2,637$)

17. Type of effect size reported

$F(1, 5575) = 32.52$, $p = .000$, $\eta^2 = .006$

$\overline{C}_{r} = 8.89$ ($Mdn = 5.48$, $SD = 10.37$, $n = 5,021$)

$\overline{C}_{d} = 4.68$ ($Mdn = 6.11$, $SD = 3.26$, $n = 556$)

18. Was a confidence interval reported?

$F(2, 5468) = 2.34$, $p = .071$, $\eta^2 = .001$

$\overline{C}_{\text{yes, 95\%}} = 9.21$ ($Mdn = 6.44$, $SD = 9.93$, $n = 2,621$)

$\overline{C}_{\text{no}} = 8.21$ ($Mdn = 4.63$, $SD = 10.76$, $n = 2,444$)

$\overline{C}_{\text{yes, 90\%}} = 5.79$ ($Mdn = 6.11$, $SD = 3.33$, $n = 406$)

19. Was a credibility interval reported?

$F(3, 5451) = 64.69$, $p = .000$, $\eta^2 = .034$

$\overline{C}_{\text{no}} = 7.65$ ($Mdn = 5.71$, $SD = 8.19$, $n = 3,548$)

$\overline{C}_{\text{yes, 90\%}} = 12.15$ ($Mdn = 9.33$, $SD = 17.04$, $n = 714$)

$\overline{C}_{\text{yes, 95\%}} = 10.91$ ($Mdn = 7.56$, $SD = 9.14$, $n = 594$)

$\overline{C}_{\text{yes, 80\%}} = 5.16$ ($Mdn = 2.50$, $SD = 6.27$, $n = 599$)

20. Was a fail-safe statistic reported?

$F(1, 5567) = 2.17$, $p = .141$, $\eta^2 = .000$

$\overline{C}_{\text{no}} = 8.50$ ($Mdn = 6.00$, $SD = 10.10$, $n = 5,173$)

$\overline{C}_{\text{yes}} = 7.74$ ($Mdn = 4.00$, $SD = 7.97$, $n = 396$)

21. Was there any other test of possible publication bias?

$F(1, 5407) = 46.47$, $p = .000$, $\eta^2 = .009$

$\overline{C}_{\text{no}} = 8.80$ ($Mdn = 6.11$, $SD = 10.11$, $n = 5,296$)

$\overline{C}_{\text{yes}} = 2.31$ ($Mdn = 2.67$, $SD = 2.04$, $n = 113$)

Note: $\overline{C}$ = mean number of citations per year. $N = 5,581$ meta-analytically derived effect-size estimates; n = number of effect size estimates (out of the total of 5,581) for which information was available to code for each of the 21 choices and judgment calls.

of citations per year as the dependent variable. Identical to the ANCOVAs using effect sizes as the dependent variable, the covariates were 11 dummy-coded variables representing the 12 topic areas described in the Method section. Compared with Table 2, which shows results regarding meta-analytically derived effect sizes as the dependent variable, the mean η^2 for the values shown in Table 4 is .04 (vs. .007 based on Table 2) and the median η^2 is .017 (vs. .004 based on Table 2). A formal test confirmed that indeed the mean η^2 across the 21 methodological choices and judgment calls explaining variance in citations per year is larger than the mean η^2 explaining variance in the meta-analytically derived effect sizes, $F(1, 40) = 4.65$, $p < .05$, $\eta^2 = .10$. In short, the amount of variance explained by choices and judgment calls is almost six times larger when explaining citations per year compared with effect sizes.

An examination of the citation rates across the various choices and judgment calls included in Table 4 indicates that the most widely cited meta-analyses have made methodological choices and judgment calls that are consistent with the psychometric approach to meta-analysis (i.e., Hunter & Schmidt, 2004). In terms of design, if studies are eliminated from the meta-analytic database, the most widely cited meta-analyses base this decision primarily on missing information, and the list of studies used in the meta-analysis is not included in the article. Second, in terms of analysis, most widely cited meta-analyses adopt a random-effects model, use the Hunter and Schmidt (1990, 2004) procedures and the Hunter-Schmidt software package, correct for range restriction (multiple corrections if available) in both the independent and the dependent variable, and correct for measurement error in the independent variable. In terms of reporting results, studies receive more citations if they report rs (vs. ds) and if the r is reported together with a 90% credibility interval. Finally, regarding robustness and replicability of results, the most widely cited meta-analyses do not report a fail-safe statistic or any other test of possible publication bias. In short, methodological choices and judgment calls are associated with the extent to which a meta-analysis is cited, which is an important piece of evidence regarding an article's impact and influence in the field (Aguinis, Pierce, et al., 2009; Podsakoff et al., 2008).

Meta-Analytic Objective: Theory Building and Theory Testing

Thus far, our results have revealed that methodological choices and judgment calls are not as strongly related to the size of meta-analytically derived effect sizes compared to an article's impact on the field (i.e., number of citations). In this section, we report results regarding relationships among the main goal of a meta-analysis (i.e., theory building vs. theory testing) and the resulting effect sizes. In addition, we examine the relationship between the theory-building and theory-testing objectives with the extent to which a meta-analysis has had scholarly impact, as indicated by the number of citations per year. Figure 3 shows trends in theory building and theory testing from 1982 to 2009. As is shown in this figure, meta-analysis has been used increasingly to both build new theories and test existing theories. Accordingly, given this longitudinal trend, our model also includes an article's year of publication.

Table 5 includes zero-order correlations among the theory-building and theory-testing ratings, together with the year of publication for each article, each meta-analytically derived effect size, and the number of citations received per year by each article. We tested a model

Figure 3
Trends in Theory Building and Theory Testing in Published
Meta-Analyses (1982-2009)
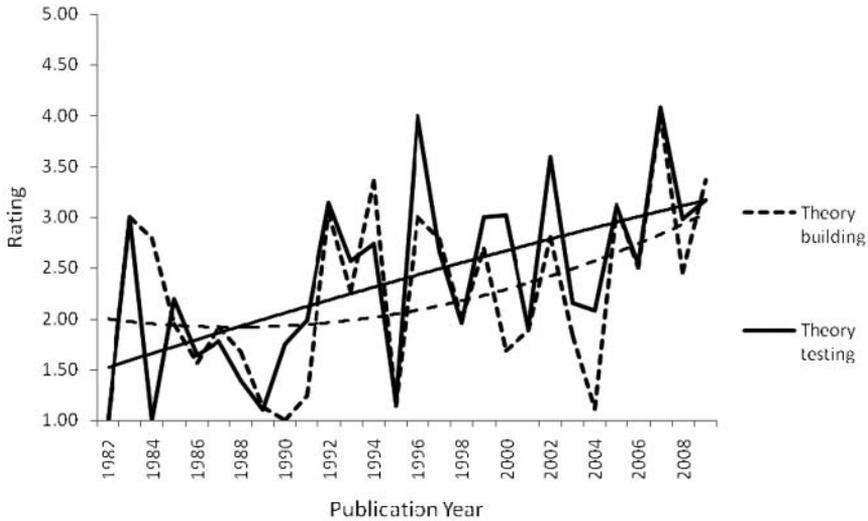
Note: Rating indicates the scales used to assess the extent to which the goal of each published meta-analysis is to (a) build theory and (b) test theory (from 1 = *low* to 5 = *high*).

Table 5
Descriptive Statistics and Zero-Order Correlations

Variable	*M*	*SD*	1	2	3	4	5
1. Theory building	2.59	1.30					
2. Theory testing	2.93	1.28	.81**				
3. Year of publication	2,001.34	6.12	.25**	.24**			
4. Citations per year	8.47	9.97	−.20**	−.14**	−.27**		
5. |*r*|	.26	.18	.02	−.00	−.04**	.03*	

Note: N = 5,581 meta-analytically derived effect-size estimates. Year of publication ranges from 1982 to 2009.
*p < .05. **p < .01.

of the relationship among variables included in Table 5 based on the following reasoning. First, the more a study purposefully tests an existing theory and the more it builds a new theory, the greater should be its impact on the field (i.e., higher citation rates per year). Second, year of publication should be positively related to theory testing and theory building because, as the field progresses, there should be more emphasis on both theory testing and theory building (which seems to be the case based on the graph included in Figure 3). Third,

theory building and theory testing should be positively related because, as noted by Bacharach (1989), building versus testing is a false dichotomy: "If it is not testable, no matter how profound or aesthetically pleasing it may be, it is not a theory" (p. 512). In fact, Colquitt and Zapata-Phelan (2007) found a positive relationship between theory building and theory testing. In addition, if the field of management is indeed making theoretical progress, theory building and theory testing should be positively related to the resulting effect size. In other words, the more a study purposefully tests an existing theory or builds a new theory (i.e., higher scores on the scales described in the Method section), the larger should be the resulting meta-analytically derived effect size. Finally, we anticipated that effect size would be positively related to citations per year. In other words, meta-analyses reporting larger effect sizes, indicating that we are able to predict outcomes of interest with greater precision, should be cited more frequently than meta-analyses reporting smaller effect sizes. This would suggest that, collectively, researchers pay more attention to more mature theories compared with more nascent theories that are not as effective at predicting outcomes of interest (i.e., a universalistic as opposed to a particularistic state of our science).

Our resulting model is shown in Figure 4. We tested the fit of this model using Amos 17.0.0 with raw data as input and maximum likelihood estimation. Results support the fit of the model: comparative fit index (CFI) = .99, incremental fit index (IFI) = .99, normed fit index (NFI) = .99, root mean square error of approximation (RMSEA) = .04, and $\chi^2(2, N = 5,581) = 11.24$, $p = .001$. Results suggest excellent fit based on the CFI, IFI, NFI, and RMSEA. However, the χ^2 for the model is statistically significant, but this is expected given our large $N = 5,581$ (Aguinis & Harden, 2009).

Figure 4 shows that, as expected, theory testing has a positive relationship with citations per year ($\beta = .10$, $p < .01$), but theory building has a negative relationship with citations per year ($\beta = -.23$, $p < .01$). Second, also as expected, there is a positive relationship between theory building and year of publication ($\beta = .25$, $p < .01$) and theory testing and year of publication ($\beta = .24$, $p < .01$), and a positive relationship between theory building and theory testing ($\beta = .81$, $p < .01$). Third, as expected, there is a positive relationship between theory building and effect size ($\beta = .06$, $p < .01$) but a negative relationship between theory testing and effect size ($\beta = -.05$, $p < .05$). Finally, the relationship between effect size and citations per year is statistically nonsignificant ($\beta = .02$, $p = .09$). In sum, the model has excellent fit, and the results regarding six of the nine path coefficients were as expected, but the following three were not: The negative relationship between theory building and citations per year, the negative relationship between theory testing and effect size, and the statistically nonsignificant relationship between effect size and citations per year.

Discussion

Although not one of our main goals, we collected new evidence in support of the assertion that meta-analysis is one of the most influential methodological tools in management and related fields. Specifically, the mean number of total citations for the 196 meta-analyses we reviewed is 88.44 (*Mdn* = 45.00), which is almost three times as large as the mean number of total citations of 31.45 received by the 667 empirical articles published in *AMJ* from 1963

Figure 4
Model Showing the Relationships Among Theory Building, Theory Testing, Effect Size (i.e., |r|), Year of Publication, and Web of Science Citations Per Year
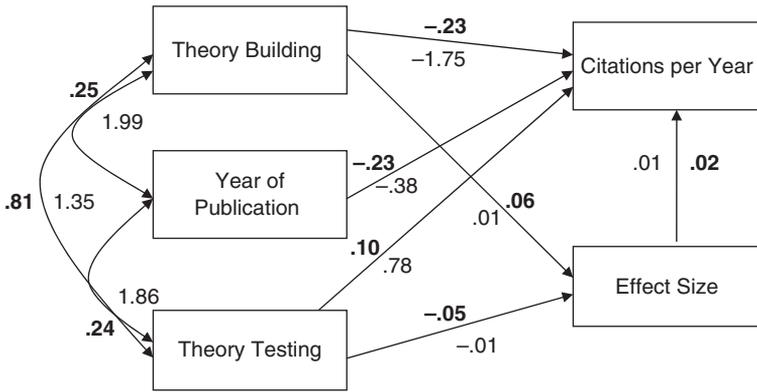
Note: Standardized coefficients are in bold type, and unstandardized coefficients are in regular type. Straight arrows imply direct relationships, and curved arrows imply covariances. For all path coefficients, *p* < .01, except for the paths from theory testing to effect size (*p* < .05) and from effect size to citations per year (*p* > .05).

to 2007 (Colquitt & Zapata-Phelan, 2007). Also, note that Colquitt and Zapata-Phelan included all empirical articles in *AMJ,* including all meta-analyses as well, and their review includes articles published as far back as 1963, whereas the earliest articles included in our review date back to the early 1980s. This difference gives an advantage to older articles because they have more time to accumulate citations. Thus, the relative advantage of meta-analyses compared with primary-level studies in terms of impact on the field is in all likelihood even greater than we estimated.

Dozens of articles have been published and continue to be published regularly on meta-analytic procedures. Supposedly, these methodological refinements should enhance the usefulness of our theories. Because a useful theory is able to explain and predict variance in relevant outcomes, dependent variables, or criteria, an indicator of a theory's practical usefulness is whether meta-analytically derived effect sizes vary as a function of the implementation of various methodological choices available to researchers. In addition to methodological refinements, several authors have written about the presumed dangers of a variety of judgment calls that researchers engage in while conducting a meta-analysis. Supposedly, the implementation of these judgment calls also affects the obtained effect sizes. Certain judgment calls, which are not aligned with what may be considered "best practices," should lead to more errors in prediction and smaller effect sizes compared with judgment calls better aligned with "best practices" (Bullock & Svyantek, 1985; Wanous et al., 1989). Many other articles have been published elsewhere issuing warnings regarding how judgment calls have important effects on meta-analytic results (e.g., Algera, Jansen, Roe, & Vijn, 1984;

Bobko & Roth, 2003; Bobko & Stone-Romero, 1998; Guzzo, Jackson, & Katzell, 1987; Matt, 1989; Rothstein, 2003). As noted earlier for the case of methodological choices, the implementation of various judgment calls is assumed to have an impact on the magnitude of the meta-analytically derived effect sizes.

Our results, based on 5,581 effect sizes reported in 196 published meta-analyses, show that the 21 methodological choices and judgment calls we investigated do not explain a substantial amount of variance in the meta-analytically derived effect sizes. In fact, the mean effect size (i.e., η^2) for the impact of methodological choices and judgment calls on the magnitude of the meta-analytically derived effect sizes is only .007. The median η^2 value is an even smaller value of .004. A perusal of Table 2 shows the small impact of these presumed effects. For example, if a meta-analyst chooses to eliminate studies from the database, the resulting absolute value mean *r* is .26 (*Mdn* = .24). If no elimination takes place, the resulting mean *r* is .25 (*Mdn* = .21). If a fail-safe statistic was reported, the resulting absolute value mean *r* is .26 (*Mdn* = .23). If no fail-safe statistics are reported, the resulting mean *r* is also .26 (*Mdn* = .22). Differences shown in Table 2 between *r*s on the order of .05 or less are not sufficiently large to change our views regarding the predictive potential and practical usefulness of our theories in most contexts. In addition, specifically regarding the implementation of corrections for statistical and methodological artifacts, our results indicate that, across 138 meta-analyses reporting approximately 3,000 effect sizes, the use of corrections improves the resulting meta-analytically derived effect size by about .05 correlation coefficient units (i.e., from about .21 to about .25). Stated differently, implementing corrections for statistical and methodological artifacts improves our knowledge of variance in outcomes by about only .25%. Across the 12 topic areas we investigated, the increase in the resulting effect sizes due to correcting for statistical and methodological artifacts ranges from about .03 to .09 correlation coefficient units.

The above results lead to two important conclusions. First, regarding judgment calls, warnings about the impact of these pseudo-idiosyncratic decisions may not be fully justified. Our results suggest that, overall, meta-analytic methodology is quite robust to the types of decisions we make and there is not a substantive decrease in meta-analytically derived effect sizes when researchers make one type of call or another. This is good news and speaks positively about the robustness of meta-analysis as a methodological tool. The fact that researchers engage in a variety of idiosyncratic practices, at least the 21 we investigated in our study, does not seem to have a noticeable impact on substantive conclusions. Second, dozens of articles on meta-analytic refinements have been published and continue to be published on an ongoing basis. However, our results indicate that implementing the specific choices we studied does not affect results in a substantive manner. The magnitude of meta-analytically derived effect sizes does not increase in a substantial way based on the implementation of various choices, which means that the degree to which our theories explain and predict outcomes of interest is also not improved in a substantive way. We did find some changes in the resulting effect sizes (cf. Table 2), and this is driven in part by the large sample size and the resulting statistical power to detect even very small differences, but these differences are small in magnitude and do not affect substantive conclusions. An improvement in a meta-analytically derived *r,* for example, from .30 to .32, may seem impressive if we describe this change as a "7% improvement" in the relationship between a

predictor and an outcome of interest (e.g., Oh et al., 2008). However, this type of "improvement" means that we now explain 10% of variance in an outcome rather than 9%. Although such a small increase may be practically significant in a handful of contexts (see Cortina & Landis, 2009, for a few examples), it is hardly large enough to conclude that we now explain the outcome so much better that the practical usefulness of a theory is substantially greater.

Reasons Why Our Results Are Not Necessarily at Odds With Previous Research

Our results may seem counterintuitive and inconsistent with research results on meta-analytic methodology. However, although admittedly provocative, our results and conclusions are not as inconsistent with previously published research as they may seem at first glance.

Let us first consider the case of methodological choices. This body of research relies mainly on Monte Carlo simulations in which researchers specify population parameters, create samples with various characteristics (e.g., sample size, amount of measurement error in predictor and criterion scores), and then assess the results in relationship with the expectation (i.e., population parameters known to the researcher). There are two important characteristics that the vast majority of Monte Carlos studies have in common. First, as noted by Murphy (2003) regarding validity generalization (i.e., psychometric meta-analysis) in particular, but equally applicable to any type of meta-analytic procedure,

> a central weakness of most tests of the accuracy of VG estimates is the gap between the assumptions needed to develop and test these models and the actual process by which validity results are produced and generated. . . . Assumptions of normality, random sampling, independence, and so forth are routinely violated in most studies in the behavioral and social sciences. (p. 17)

In other words, Monte Carlo studies rely on assumptions that are almost always untenable in the conduct of an actual meta-analysis. Consequently, Monte Carlo results are likely to exaggerate differences produced by methodological refinements, and therefore, it is not clear that Monte Carlo simulations can assess the actual effects of these refinements in actual (as opposed to simulated) meta-analyses.

A second common characteristic of this body of work is that a careful examination of results derived from Monte Carlo studies suggests that the improvements due to refinements may seem to be, but actually are not, substantial in most cases. In fact, the majority of Monte Carlo studies emphasize the percentage improvement as a consequence of the introduction of a specific refinement. We do not intend to criticize this body of work, but we offer a couple of specific examples regarding this important point. One example is Aguinis and Whitehead's (1997) conclusion that "in the presence of indirect range restriction, variability across study-level rs can be underestimated by as much as 8.50%" (p. 537). Based on this result, Aguinis and Whitehead recommended that future meta-analyses correct for the effects of indirect range restriction. As a second illustration, Aguinis (2001) concluded that a new and improved estimator of the sampling variance of correlations improved the estimation in about 50% of the conditions of the simulation design and the improvement was as large as

27.88% for some conditions (p. 578). Consequently, Aguinis recommended the use of the improved sampling variance estimator. However, a close examination of these and many other additional articles describing Monte Carlo results suggests that improvements described in terms of percentages do not necessarily translate into substantial improvements if they are expressed in effect-size metric. In fact, these results are even less substantial if they are expressed in terms of theory usefulness, as indicated by the percentage of additional variance explained in the relevant outcome. Moreover, the impact of these improvements on the resulting effect sizes is similarly small in magnitude to those we obtained in our study.

We are in favor of any type of methodological improvement, because "a hallmark of a science is continuous improvement in accuracy of measurements and estimations of theoretically important values" (Aguinis, 2001, p. 587). Moreover, we agree with Schmidt and Hunter (2003), who noted that "even if estimates are quite accurate, it is always desirable to make them more accurate if possible" (p. 41). Based on our data, however, we conclude that the choices we investigated do not necessarily lead to concomitant substantial improvements in the usefulness of our theories. And, in retrospect, these findings are not at odds with the existing literature if we consider the nature of the Monte Carlo studies upon which this literature is based and the magnitude of results expressed in terms of the degree of theory improvements.

Now, let us consider the case of judgment calls. The majority of published sources about judgment calls raise issues, questions, and warnings about the potential effects of these judgment calls (e.g., Guzzo et al., 1987). A few sources provide selective examples of how these judgment calls can have an impact (e.g., Bullock & Svyantek, 1985). Also, fewer sources still actually conduct an analysis of selective published cases to illustrate how judgment calls can have an impact (e.g., Wanous et al., 1989). This body of literature has concluded that such judgment calls have the potential to affect substantive conclusions and that they "*can* produce widely divergent results" (Steiner, Lane, Dobbins, Schnur, & McConnell, 1991, p. 621, italics added).We were not subjectively selective in how we chose the meta-analyses that we included in our study. To the contrary, we extended a wide net and examined meta-analyses published in five flagship journals from 1982 to 2009, including a total of 5,581 individual effect sizes. In short, we do not believe that our results are necessarily at odds with the results of previous research regarding the impact of judgment calls, because this systematic and comprehensive research has not yet been conducted using actual published meta-analyses as we did in the present study.

If choices and judgment calls do not affect results in a substantive way, what does? We believe that in the particular case of meta-analysis, we are not likely to see substantial advancements if the improvement of methodological tools lacks concomitant good theory (Bobko & Roth, 2008), and, if there is, there are no important improvements in the design and measurement issues at the primary level (Aguinis, Pierce, et al., 2009; Van Maanen et al., 2007). The old adage "garbage in–garbage out" is most pertinent in the context of meta-analytic reviews. If the set of primary-level studies is not of high quality, methodological tweaks and refinements at the meta-analytic level will not be able to improve the quality of the results. If researchers conducting primary-level studies do not take advantage of better design, measurement, and analysis tools, it is unlikely that the usefulness of theories built or tested using meta-analysis will increase dramatically (Aguinis, Pierce, Bosco, Dalton, & Dalton, in press).

Meta-Analytic Choices and Judgment Calls, Theory Building and Testing, Obtained Effect Sizes, and Scholarly Impact

Our results suggest that, over time, there is a trend for meta-analyses to increasingly attempt to make contributions to both building new theory and testing existing theories (see Figure 3). In fact, Figure 4 shows that these relationships are quite strong. Specifically, the standardized and unstandardized coefficients are 0.25 and 1.99 between theory building and year of publication and 0.24 and 1.86 between theory testing and year of publication. This is a similar trend to that found by Colquitt and Zapata-Phelan (2007). However, in contrast to Colquitt and Zapata-Phelan, and because our study includes information about effect sizes and theirs did not, we were actually able to examine the relative success in terms of scholarly impact of theory-building versus theory-testing approaches. Our results suggest that meta-analyses that attempt to make mainly a contribution to theory testing are more successful than meta-analyses that attempt to make mainly a contribution to theory building. In fact, the more a meta-analysis attempts to test an existing theory, the more frequently the meta-analysis is cited. In contrast, the more a meta-analysis attempts to build new theory, the lower the number of citations it receives. Results also suggest a positive relationship between theory building and absolute magnitude of effect sizes. In short, meta-analyses that build new theories are more likely to obtain larger effect sizes but are less likely to be cited compared with meta-analyses that test existing theories.

There is a likely reason why theory building is positively related to the size of the obtained effects whereas theory testing is negatively related to effect size. Theory building is usually an inductive and exploratory process, whereas theory testing is usually a deductive and confirmatory process. Exploratory analyses rely on chance more so than confirmatory analyses do. Consequently, compared with confirmatory analyses, exploratory analyses usually yield overestimates of effects. It has been suggested that exploratory analyses (e.g., factor analysis, validation research) be cross-validated by further research because estimates are usually inflated (e.g., Bandalos & Boehm-Kaufman, 2009; Cascio & Aguinis, 2005; Conway & Huffcutt, 2003).

Also, results based on testing the model in Figure 4 indicate that the more a meta-analysis attempts to test an existing theory, the larger the number of citations (standardized coefficient of .10), whereas the more a meta-analysis attempts to build new theory, the lower the number of citations (standardized coefficient of –.23). These results provide evidence that the introduction of new theories and paradigms is very difficult and that researchers, like people in general, are not prone to changing their ideas easily (Kuhn, 1962). A similar conclusion was drawn based on a review of articles focusing on changes regarding methodological practices in the organizational sciences (e.g., Aguinis, Pierce, et al., 2009). In other words, researchers are more likely to cite meta-analyses that address mainly established, and known, theories rather than those that attempt to build new, unknown theories.

In contrast to our expectation regarding universalism, our results provide support for a competing scientific particularism perspective given the statistically nonsignificant relationship between effect size and citations per year (i.e., $\beta = .01$, $p > .05$). In other words, meta-analyses that report larger effect sizes and, hence, explain more variance in relevant outcomes do not receive more attention from the scholarly community compared with meta-analyses

that report smaller effect sizes. It seems that what researchers consider sufficiently interesting to be cited does not necessarily coincide with those results that have greater ability to explain outcomes of interest. These results are consistent with the widely documented science–practice divide in management and related fields: Researchers choose to study what are interesting phenomena to them, but these phenomena, and the obtained results, may not be relevant to practitioners (Cascio & Aguinis, 2008), or the obtained effects may be so weak that they are not likely to lead to noticeable results for practice (Combs, 2010).

Taken together, these results lead to interesting implications in terms of a meta-analyst's dual goals to improve the usefulness of a theory and to publish an article that will be widely cited. First, choosing to conduct a meta-analysis and making procedural choices consistent with the Hunter-Schmidt approach is likely to enhance the chances that the article will be accepted for publication. Table 1 indicates that the majority of meta-analyses follow this approach (i.e., more than 83% of effect sizes were computed using the Hunter-Schmidt procedures). Note that this conclusion assumes that submitted manuscripts using various approaches are not accepted at drastically different rates. In fact, the reversal of the conclusion that the Hunter-Schmidt procedure is the most popular would necessitate that, for example, meta-analyses using the Hedges-Olkin approach be rejected at a rate approximately 20 times larger compared with meta-analyses using the Hunter-Schmidt approach. Second, also making choices consistent with the Hunter-Schmidt approach is likely to enhance the article's subsequent citations (see Table 4). Note that the small base rate for methods other than Hunter-Schmidt's actually makes it more difficult to find a relationship between type of method and citations. This is because the variance of a categorical variable such as type of method decreases as the sample sizes across type of method diverge, and the smaller the variance, the more difficult it is to find relationships with other variables. Specifically,

the variance of the categorical variable type of method Z is $S^2 = \dfrac{\sum (Z_i - \bar{Z})^2}{N-1} = \dfrac{Np(1-p)}{N-1}$

(Aguinis, Boik, & Pierce, 2001, p. 320). If, for the sake of simplicity, we assume that Z has two categories only, then $N = n_1 + n_2$ and $p = n_1/N$, where n_1 is the sample size in one category and n_2 is the sample size in the second category. So, the further the value of p from .50 (i.e., $p = .50$ means equal sample size across categories), the smaller the variance of Z. Third, although choosing the Hunter-Schmidt approach is likely to enhance the chances the article will be published and cited widely, this choice does not necessarily result in larger effect sizes, which indicate a theory's ability to predict relevant outcomes and its practical usefulness (see Table 2). Moreover, reporting larger meta-analytically derived effect sizes does not necessarily lead to a higher citation rate. Finally, attempting to develop new theory as opposed to test an existing theory is likely to lead to larger effect sizes. However, this choice will be made at the expense of scholarly impact (i.e., lower expected citations; see Figure 4). In short, choosing to implement the Hunter-Schmidt approach increases the chances that the article will be published and cited but is not necessarily related to the size of the obtained effect. In addition, attempting to build new theory is likely to result in larger effects than testing an existing theory, but a theory-building attempt is likely to lead to less impact on the academic community (i.e., fewer citations).

It seems that the Hunter-Schmidt approach is dominant in terms of what is considered best meta-analytic practice. Journals and scientific fields have cultures with underlying

norms about accepted practices (cf. Colquitt & Zapata-Phelan, 2007) and our results indicate that the Hunter-Schmidt approach is the zeitgeist in management. However, this zeitgeist is not shared by other fields (Aguinis, Gottfredson, et al., in press). Specifically, Schmidt, Oh, and Hayes (2009) reviewed meta-analyses published from 1978 to 2006 in *Psychological Bulletin,* which is a journal devoted to publishing comprehensive scientific psychology research integrations that are often cross-disciplinary and broad in nature. Their review revealed that only 11.80% of articles used the Hunter-Schmidt approach. Moreover, our results indicate that adopting this approach is not necessarily improving the obtained effect sizes, compared with the use of alternative procedures. Looking toward the future, it is likely that the primacy of the Hunter-Schmidt approach will remain unchallenged in management, given that several reviews of methodological usage have documented that, once they are established as dominant, there is a "scientific community's persistence in the use of particular methods" (Podsakoff & Dalton, 1987, p. 433; see also Aguinis, Pierce, et al., 2009).

Potential Limitations and Directions for Future Research

In this section, we describe four potential limitations of our study, together with suggested directions for future research. These potential limitations are, for the most part, related to the scope of our review.

First, an important strength of our review is its inclusiveness. Our study included a very large and diverse number of effect sizes assessing relationships in many management subfields. For example, among the 5,581 effect sizes we analyzed, we investigated the relationship between personality and performance, job satisfaction and organizational citizenship behaviors, mentoring and career-related outcomes, network structure and team performance, CEO duality and firm performance, and strategic resources and firm performance, just to name a few. The other side of the coin of the inclusiveness of our study is that, as noted by an anonymous reviewer, the majority of our analyses involved across-study comparisons with the underlying assumption that methodological choices and judgment calls have been "randomly assigned" to the published meta-analyses. By coding and controlling for topic area, we were able to minimize the biasing effect of one potentially important confound, but not all. Accordingly, our results suggest that, overall, methodological choices and judgment calls have little impact on effect sizes, and this conclusion is reinforced by results of the within-study comparisons regarding the impact of corrections for statistical and methodological artifacts (cf. Table 3). However, we emphasize that our results do not suggest that such choices do not have an important impact on effect sizes in all circumstances.

Second, we conducted a content analysis including five journals. Our decision was guided by the fact that *AMJ, JAP, JOM, PPsych,* and *SMJ* are five of the most prestigious journals in the field (Podsakoff et al., 2008). Given the amount of data we were able to collect, we are doubtful that the inclusion of additional journals would alter our substantive conclusions. In fact, our initial investigation included *JAP* and *PPsych* only (Aguinis, Dalton, Bosco, Pierce, & Dalton, 2009). The addition of *AMJ, JOM,* and *SMJ* to our database did not alter our conclusions in a substantive manner.

Third, we investigated the impact of 21 methodological choices and judgment calls. Readers may be able to identify additional methodological choices and judgment calls that can

be made in the course of conducting a meta-analysis. However, we knew we would be able to obtain necessary data from the published studies regarding the issues we examined, and that guided our choice, given difficulties in obtaining data directly from authors. For example, Aguinis et al. (2005) also conducted a review of effect sizes (in the context of interaction effects in multiple regression) and reported that fewer than 5% of the authors contacted directly provided data not already available in their published articles. Given the number of choices and calls we examined and the consistency of the results, we are doubtful that examining additional choices and calls would change our study's conclusions in a substantial manner.

Finally, our study focused on the relationship between choices and judgment calls and their relationship to summary effect sizes. However, methodological choices and judgment calls may explain variability in across-study variation in effect sizes across meta-analyses as well as the presence and magnitude of moderating effects. Meta-analysis is not only concerned with the summary effect sizes but also interested in the extent to which effect sizes vary across studies and the potential moderator variables that may explain this variance (Aguinis, Pierce, et al., in press). Thus, future research could investigate the extent to which choices and judgment calls are related to across-study variance and substantive conclusions regarding the presence and magnitude of moderating as well as mediating effects.

Conclusions

Our results indicate that, overall, the 21 methodological choices available and judgment calls involved in the conduct of a meta-analysis that we investigated have little impact on substantive conclusions. Thus, the present study based on actual meta-analyses casts doubt on previous warnings that judgment calls have an important and possible detrimental impact on substantive conclusions. Results also indicate that methodological choices and judgment calls are related to the impact of (i.e., number of citations received by) a meta-analysis, but these choices and judgment calls are not as strongly related to the size of the obtained effect, which is an indicator of a theory's predictive potential and practical usefulness, and larger effect sizes do not lead to higher citation rates. Taken together, our results provide a comprehensive data-based understanding of how meta-analytic reviews are conducted and the implications of these practices for theory building and testing, obtained effect sizes (i.e., predictive potential and practical usefulness of a theory), and scholarly impact.

References

Aguinis, H. 2001. Estimation of sampling variance of correlations in meta-analysis. *Personnel Psychology,* 54: 569-590.

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. 2005. Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology,* 90: 94-107.

Aguinis, H., Boik, R. J., & Pierce, C. A. 2001. A generalized solution for approximating the power to detect effects of categorical moderator variables using multiple regression. *Organizational Research Methods,* 4: 291-323.

Aguinis, H., Culpepper, S. A., & Pierce, C. A. in press. Revival of test bias research in preemployment testing. *Journal of Applied Psychology*.

Aguinis, H., Dalton, D. A., Bosco, F. A., Pierce, C. A., & Dalton, C. M. 2009. *Meta-analytic choices and judgment calls: Implications for theory and scholarly impact*. Paper presented at the meeting of the Academy of Management, Chicago, IL, August.

Aguinis, H., Gottfredson, R. K., & Wright, T. A. in press. Best-practice recommendations for estimating interaction effects using meta-analysis. *Journal of Organizational Behavior*.

Aguinis, H., & Harden, E. E. 2009. Cautionary note on conveniently dismissing χ^2 goodness-of-fit test results: Implications for strategic management research. In D. D. Bergh & D. J. Ketchen (Eds.), *Research methodology in strategy and management* (Vol. 5): 111-120. Howard House, UK: Emerald Group.

Aguinis, H., & Pierce, C. A. 1998. Testing moderator variable hypotheses meta-analytically. *Journal of Management,* 24: 577-592.

Aguinis, H., Pierce, C. A., Bosco, F. A., Dalton, D. R., & Dalton, C. M. in press. Debunking myths and urban legends about meta-analysis. *Organizational Research Methods*.

Aguinis, H., Pierce, C. A., Bosco, F. A., & Muslin, I. S. 2009. First decade of *Organizational Research Methods:* Trends in design, measurement, and data-analysis topics. *Organizational Research Methods,* 12: 69-112.

Aguinis, H., Sturman, M. C., & Pierce, C. A. 2008. Comparison of three meta-analytic procedures for estimating moderating effects of categorical variables. *Organizational Research Methods,* 11: 9-34.

Aguinis, H., & Whitehead, R. 1997. Sampling variance in the correlation coefficient under indirect range restriction: Implications for validity generalization. *Journal of Applied Psychology,* 82: 528-538.

Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. 2010. Customer-centric science: Reporting research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods*, 13: 515-539.

Algera, J. A., Jansen, P. G., Roe, R. A., & Vijn, P. 1984. Validity generalization: Some critical remarks on the Schmidt-Hunter procedure. *Journal of Occupational Psychology,* 57: 197-210.

Bacharach, S. B. 1989. Organizational theories: Some criteria for evaluation. *Academy of Management Review,* 14: 496-515.

Bandalos, D. L., & Boehm-Kaufman, M. R. 2009. Four common misconceptions in exploratory factor analysis. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences:* 61-87. New York: Routledge.

Bays, C. L. 2001. Quality of life of stroke survivors: A research synthesis. *Journal of Neuroscience Nursing,* 33: 310-316.

Beal, B. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. 2003. Cohesion and performance in groups: A meta-analytic clarification of construct relationships. *Journal of Applied Psychology,* 88: 989-1004.

Bierstedt, R. 1959. Nominal and real definitions in sociological theory. In L. Gross (Ed.), *Symposium on sociological theory:* 121-144. New York: Harper & Row.

Bobko, P., & Roth, P. L. 2003. Meta-analysis and validity generalization as research tools: Issues of sample bias and degrees of mis-specification. In K. R. Murphy (Ed.), *Validity generalization: A critical review:* 67-90. Mahwah, NJ: Lawrence Erlbaum.

Bobko, P., & Roth, P. L. 2008. Psychometric accuracy and (the continuing need for) quality thinking in meta-analysis. *Organizational Research Methods,* 11: 114-126.

Bobko, P., & Stone-Romero, E. F. 1998. Meta-analysis may be another useful research tool, but it is not a panacea. *Research in personnel and human resources management* (Vol. 16): 359-397. Stamford, CT: JAI.

Boyd, B. K., Finkelstein, S., & Gove, S. 2005. How advanced is the strategy paradigm? The role of particularism and universalism in shaping research outcomes. *Strategic Management Journal,* 26: 841-854.

Bullock, R. J., & Svyantek, D. J. 1985. Analyzing meta-analysis: Potential problems, an unsuccessful replication, and evaluation criteria. *Journal of Applied Psychology,* 70: 108-115.

Callender, J. C., & Osburn, H. G. 1988. Unbiased estimation of sampling variance of correlations. *Journal of Applied Psychology,* 73: 312-315.

Cascio, W. F., & Aguinis, H. 2005. Test development and use: New twists on old questions. *Human Resource Management,* 44: 219-235.

Cascio, W. F., & Aguinis, H. 2008. Research in industrial and organizational psychology from 1963 to 2007: Changes, choices, and trends. *Journal of Applied Psychology,* 93: 1062-1081.

Colquitt, J. A., LePine, J. A., & Noe, R. A. 2000. Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology,* 85: 678-707.

Colquitt, J. A., & Zapata-Phelan, C. P. 2007. Trends in theory building and theory testing: A five-decade study of *Academy of Management Journal. Academy of Management Journal,* 50: 1281-1303.

Combs, J. G. 2010. Big samples and small effects: Let's not trade relevance and rigor for power. *Academy of Management Journal,* 53: 9-13.

Conway, J. M., & Huffcutt, A. I. 2003. A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods,* 6: 147-168.

Cooper, H. M. 1982. Scientific guidelines for conducting integrative research reviews. *Review of Educational Research,* 52: 291-302.

Cortina, J. M., & Landis, R. S. 2009. When small effect sizes tell a big story, and when large effect sizes don't. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences:* 287-308. New York: Routledge.

Dalton, D. R., & Dalton, C. M. 2008. Meta-analyses: Some very good steps toward a bit longer journey. *Organizational Research Methods,* 11: 127-147.

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. 2007. A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods,* 10: 5-34.

Erez, A., Bloom, M. C., & Wells, M. T. 1996. Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology,* 49: 275-306.

Geyskens, I., Krishnan, R., Steenkamp, J. E. M., & Cunha, P. V. 2009. A review and evaluation of meta-analysis practices in management research. *Journal of Management,* 35: 393-419.

Guzzo, R. A., Jackson, S. E., & Katzell, R. A. 1987. Meta-analysis analysis. *Research in Organizational Behavior,* 9: 407-442.

Hedges, L. V. 1982. Estimation of effect sizes from a series of experiments. *Psychological Bulletin,* 92: 490-499.

Hedges, L. V., & Olkin, I. 1985. *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

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., & Jackson, G. B. 1982. *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, 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.

James, L. R., Demaree, R. G., & Mulaik, S. A. 1986. A note on validity generalization procedures. *Journal of Applied Psychology,* 71: 440-450.

James, L. R., Demaree, R. G., Mulaik, S. A., & Ladd, R. T. 1992. Validity generalization in the context of situational models. *Journal of Applied Psychology,* 77: 3-14.

Johnson, B. T., Mullen, B., & Salas, E. 1995. Comparison of three major meta-analytic approaches. *Journal of Applied Psychology,* 80: 94-106.

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. L. 2007. Self-efficacy and work-related performance: The integral role of individual differences. *Journal of Applied Psychology,* 92: 107-127.

Kisamore, J. L., & Brannick, M. T. 2008. An illustration of the consequences of meta-analysis model choice. *Organizational Research Methods,* 11: 35-53.

Kuhn, T. S. 1962. *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kulinskaya, E., Morgenthaler, S., & Staudte, R. G. 2008. *Meta analysis: A guide to calibrating and combining statistical evidence*. West Sussex, UK: John Wiley.

Law, K. S., Schmidt, F. L., & Hunter, J. E. 1994. A test of two refinements in procedures for meta-analysis. *Journal of Applied Psychology,* 79: 978-986.

Long, J. S., & Fox, M. F. 1995. Scientific careers: Universalism and particularism. *Annual Review of Sociology,* 21: 45-71.

Matt, G. E. 1989. Decision rules for selecting effect sizes in meta-analysis: A review and reanalysis of psychotherapy outcome studies. *Psychological Bulletin,* 105: 106-115.

McDaniel, M. A., Rothstein, H. R. & Whetzel, D. L. 2006. Publication bias: A case study of four test vendors. *Personnel Psychology,* 59: 927-953.

Mill, J. S. 1843/2002. *A system of logic*. Honolulu, HI: University Press of the Pacific.

Murphy, K. R. 2003. The logic of validity generalization. In K. R. Murphy (Ed.), *Validity generalization: A critical review:* 1-29. Mahwah, NJ: Lawrence Erlbaum.

Murphy, K. R., & Newman, D. A. 2003. The past, the present, and future of validity generalization. In K. R. Murphy (Ed.), *Validity generalization: A critical review:* 403-424. Mahwah, NJ: Lawrence Erlbaum.

Newman, D. A., Jacobs, R. R., & Bartram, D. 2007. Choosing the best method for local validity estimation: Relative accuracy of meta-analysis versus a local study versus Bayes-analysis. *Journal of Applied Psychology,* 92: 1394-1413.

Oh, I., Schmidt, F. L., Shaffer, J. A., & Le, H. 2008. The Graduate Management Admission Test (GMAT) is even more valid than we thought: A new development in meta-analysis and its implications for the validity of the GMAT. *Academy of Management Learning and Education,* 7: 563-570.

Osburn, H. G., & Callender, J. 1992. A note on the sampling variance of the mean uncorrected correlation in meta-analysis and validity generalization. *Journal of Applied Psychology,* 77: 115-122.

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: 164-178.

Pierce, C. A., Block, R. A., & Aguinis, H. 2004. Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement,* 64: 916-924.

Podsakoff, P. M., & Dalton, D. R. 1987. Research methodology in organizational studies. *Journal of Management,* 13: 419-441.

Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Bachrach, D. G. 2008. Scholarly influence in the field of management: A bibliometric analysis of university and author impact on the management literature during the past quarter century. *Journal of Management,* 34: 641-720.

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

Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. *Psychological Bulletin,* 86: 638-641.

Rosenthal, R., & Rubin, D. B. 1982. Comparing effect sizes of independent studies. *Psychological Bulletin,* 92: 500-504.

Rothstein, H. R. 2003. Progress is our most important product: Contributions of validity generalization and meta-analysis to the development and communication of knowledge in I/O psychology. In K. R. Murphy (Ed.), *Validity generalization: A critical review:* 115-154. Mahwah, NJ: Lawrence Erlbaum.

Schmidt, F. L. 1992. What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist,* 47: 1173-1181.

Schmidt, F. L. 1996. Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods,* 1: 115-129.

Schmidt, F. L. 2008. Meta-analysis: A Constantly evolving research integration tool. *Organizational Research Methods,* 11: 96-113.

Schmidt, F. L., & Hunter, J. E. 2003. History, development, evolution, and impact of validity generalization and meta-analysis methods, 1975-2001. In K. R. Murphy (Ed.), *Validity generalization: A critical review:* 31-66. Mahwah, NJ: Lawrence Erlbaum.

Schmidt, F. L., Oh, I., & Hayes, T. L. 2009. Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology,* 62: 97-128.

Schmidt, F. L., Pearlman, K., & Hunter, J. E. 1980. The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. *Journal of Applied Psychology,* 33: 705-724.

Schmidt, F. L., & Raju, N. 2007. Updating meta-analytic research findings: Bayesian approaches versus the medical model. *Journal of Applied Psychology,* 92: 297-308.

Schmidt, F. L., Shaffer, J. A., & Oh, I. 2008. Increased accuracy for range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology,* 61: 827-868.

Schulze, R., Holling, H., & Bohning, D. 2003. *Meta-analysis: New developments and applications in medical and social sciences.* Cambridge, MA: Hogreffe & Huber.

Steel, P. D., & Kammeyer-Mueller, J. D. 2002. Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology,* 87: 96-111.

Steiner, D. D., Lane, I. M., Dobbins, G. H., Schnur, A., & McConnell, S. 1991. A review of meta-analyses in organizational behavior and human resources management: An empirical assessment. *Educational and Psychological Measurement,* 51: 609-626.

Tait, M., Padgett, M. Y., & Baldwin, T. 1989. Job and life satisfaction: A reevaluation of the strength of the relationship and gender effects as a function of the date of the study. *Journal of Applied Psychology,* 74: 502-507.

Van Maanen, J., Sørensen, J. B., & Mitchell, T. R. 2007. Introduction to special topic forum: The interplay between theory and method. *Academy of Management Review,* 32: 1145-1154.

Wanous, J. P., Sullivan, S. E., & Malinak, J. 1989. The role of judgment calls in meta-analysis. *Journal of Applied Psychology,* 74: 259-264.