

## Sage Research Methods

# Methods of Meta-Analysis: Correcting Error and Bias in Research Findings

For the most optimal reading experience we recommend using our website.

<https://methods.sagepub.com/book/mono/methods-of-meta-analysis-3e/toc>

**Author:** Frank L. Schmidt, John E. Hunter

**Pub. Date:** 2017

**Product:** Sage Research Methods

**DOI:** <https://doi.org/10.4135/9781483398105>

**Methods:** Correlation, Meta-analysis, Effect size

**Keywords:** population, estimates, errors, job performance, corrections, job satisfaction, moderator

**Disciplines:** Anthropology, Business and Management, Criminology and Criminal Justice, Communication and Media Studies, Counseling and Psychotherapy, Economics, Education, Geography, Health, History, Marketing, Nursing, Political Science and International Relations, Psychology, Social Policy and Public Policy, Social Work, Sociology, Science, Technology, Computer Science, Mathematics, Medicine

**Access Date:** September 16, 2025

**Publisher:** SAGE Publications, Ltd

**City:** 55 City Road

**Online ISBN:** 9781483398105

© 2017 SAGE Publications, Ltd All Rights Reserved.

## Front Matter

- [Copyright](#)
- [List of Tables](#)
- [List of Figures](#)
- [Preface](#)
- [Acknowledgments](#)
- [About the Authors](#)

## Chapters

- [Part 1 | : Introduction to Meta-Analysis](#)
  - [Chapter 1 | Integrating Research Findings across Studies](#)
  - [Chapter 2 | Study Artifacts and their Impact on Study Outcomes](#)
- [Part 2 | : Meta-Analysis of Correlations](#)
  - [Chapter 3 | Meta-Analysis of Correlations Corrected Individually for Artifacts](#)
  - [Chapter 4 | Meta-Analysis of Correlations Using Artifact Distributions](#)
  - [Chapter 5 | Technical Questions in Meta-Analysis of Correlations](#)
- [Part 3 | : Meta-Analysis of Experimental Effects and other Dichotomous Comparisons](#)
  - [Chapter 6 | Treatment Effects: Experimental Artifacts and their Impact](#)
  - [Chapter 7 | Meta-Analysis Methods for d Values](#)
  - [Chapter 8 | Technical Questions in Meta-Analysis of d Values](#)
- [Part 4 | : General Issues in Meta-Analysis](#)
  - [Chapter 9 | General Technical Issues in Meta-Analysis](#)
  - [Chapter 10 | Cumulation of Findings within Studies](#)
  - [Chapter 11 | Different Methods of Meta-Analysis and Related Software](#)
  - [Chapter 12 | Locating, Evaluating, Selecting, and Coding Studies and Presentation of Meta-Analysis Results](#)
  - [Chapter 13 | Availability Bias, Source Bias, and Publication Bias in Meta-Analysis](#)
  - [Chapter 14 | Summary of Psychometric Meta-Analysis](#)

## Back Matter

- [Appendix Windows-Based Meta-Analysis Software Package Version 2.0](#)
- [References](#)
- [Author Index](#)

# Copyright



Los Angeles | London | New Delhi  
Singapore | Washington DC

## FOR INFORMATION:

SAGE Publications, Inc.  
2455 Teller Road  
Thousand Oaks, California 91320  
E-mail: [order@sagepub.com](mailto:order@sagepub.com)

SAGE Publications Ltd.  
1 Oliver's Yard  
55 City Road  
London EC1Y 1SP  
United Kingdom

SAGE Publications India Pvt. Ltd.  
B 1/1 Mohan Cooperative Industrial Area  
Mathura Road, New Delhi 110 044  
India

SAGE Publications Asia-Pacific Pte. Ltd.  
3 Church Street  
#10-04 Samsung Hub  
Singapore 049483

Acquisitions Editor: Vicki Knight  
Editorial Assistant: Jessica Miller  
Digital Content Editor: Katie Guarino  
Production Editor: Laura Barrett  
Copy Editor: Gillian Dickens  
Typesetter: C&M Digital (P) Ltd.  
Proofreader: Wendy Jo Dymond  
Indexer: Will Ragsdale  
Cover Designer: Candice Harman  
Marketing Manager: Nicole Elliott

Copyright © 2015 by SAGE Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Printed in the United States of America

*Library of Congress Cataloging-in-Publication Data*

Hunter, John E. (John Edward), 1939–2002.

Methods of meta-analysis : correcting error and bias in research findings / Frank L. Schmidt, University of Iowa, John E. Hunter, Michigan State University. — Third edition.

pages cm

Includes bibliographical references and index.

ISBN 978-1-4522-8689-1 (hardcover : alk. paper)

1. Social sciences—Statistical methods. 2. Meta-analysis.  
I. Schmidt, Frank L. II. Title.

HA29.H847 2015

300.72—dc23 2013039086

This book is printed on acid-free paper.

14 15 16 17 18 10 9 8 7 6 5 4 3 2 1

FOR INFORMATION:

SAGE Publications, Inc.

2455 Teller Road

Thousand Oaks, California 91320

E-mail: [order@sagepub.com](mailto:order@sagepub.com)

SAGE Publications Ltd.

1 Oliver's Yard

55 City Road

London EC1Y 1SP

United Kingdom

SAGE Publications India Pvt. Ltd.

B 1/I 1 Mohan Cooperative Industrial Area

Mathura Road, New Delhi 110 044

India

SAGE Publications Asia-Pacific Pte. Ltd.

3 Church Street

#10-04 Samsung Hub

Singapore 049483

Copyright © 2015 by SAGE Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Printed in the United States of America

*Library of Congress Cataloging-in-Publication Data*

Hunter, John E. (John Edward), 1939–2002.

Methods of meta-analysis: correcting error and bias in research findings / Frank L. Schmidt, University of Iowa, John E. Hunter, Michigan State University. —

Third edition.

pages cm

Includes bibliographical references and index.

ISBN 978-1-4522-8689-1 (hardcover: alk paper)

1. Social sciences—Statistical methods. 2. Meta-analysis.

I. Schmidt, Frank L. II. Title.

HA29.H847 2015

300.72—dc23            2013039086

This book is printed on acid-free paper.

14 15 16 17 18 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor: Vicki Knight

Editorial Assistant: Jessica Miller

Digital Content Editor: Katie Guarino

Production Editor: Laura Barrett

Copy Editor: Gillian Dickens

Typesetter: C&M Digital (P) Ltd.

Proofreader: Wendy Jo Dymond

Indexer: Will Ragsdale

Cover Designer: Candice Harman

Marketing Manager: Nicole Elliott

# List of Tables

## Chapter 1

Table 1.1 [Correlations between organizational commitment and job satisfaction](#) 5

## Chapter 2

Table 2.1 [Study artifacts that alter the value of outcome measures \(with examples from personnel selection research\)](#) 41

Table 2.2 [Nineteen studies](#) 69

Table 2.3 [Validity coefficients from the Sears study](#) 77

Table 2.4 [Expected values of the multiple R of study characteristics with study outcomes when all study characteristics correlate zero with study outcomes and with each other](#) 81

## Chapter 3

Table 3.1 [Study artifacts that alter the value of outcome measures \(with examples from personnel selection research\)](#) 89

Table 3.2 [Correlations between socioeconomic status and police brutality \(United States\)](#) 103

Table 3.3 [Correlations between socioeconomic status and police brutality \(United States and Transylvania\)](#) 106

Table 3.4 [Tibetan employment service test validities](#) 110

Table 3.5 [A meta-analysis of hypothetical personnel selection studies with indirect range restriction](#) 153

Table 3.5a [Hypothetical validity and artifact information](#) 153

Table 3.5b [Meta-analysis worksheet](#) 153

## Chapter 4

Table 4.1 [Simplified example of the interactive method](#) 174

Table 4.2 [Organizational commitment and job satisfaction \(hypothetical results\)](#) 176

Table 4.2a [Basic information](#) 176

Table 4.2b [Meta-analysis worksheet](#) 177

Table 4.3 [Sixteen hypothetical studies](#) 180

Table 4.4 [Meta-analysis of personnel selection validities \(direct range restriction\)](#) 184

Table 4.5 [Indirect range restriction: hypothetical data for an artifact distribution meta-analysis of 12 personnel selection studies](#) 190



Table 4.6a [Hypothetical meta-analysis of performance and turnover 202](#)

Table 4.6b [Worksheet for the data in Table 4.6a 203](#)

Table 4.6c [Final data for the interactive program 205](#)

## Chapter 6

Table 6.1 [Hypothetical meta-analysis data for the effect of language intensity on persuasion \(results ordered by magnitude\) 251](#)

Table 6.2 [Power of the conventional significance test for studies with sample size  \$N = 1\$   \$N = 100\$  260](#)

Table 6.3 [Comparison ratio of corrected/actual correlations—expressed as percentages—where the corrected correlation is the estimated correlation for the continuous dependent variable computed by correcting the dichotomous variable correlation using the biserial correction formula 270](#)

## Chapter 7

Table 7.1 [Leadership training \(studies with training by outside experts\) 302](#)

Table 7.2 [Training in interpersonal skills by managers versus by experts 307](#)

Table 7.3 [Training in interpersonal skills by hours 311](#)

Table 7.4 [Bare-bones meta-analyses on hypothetical studies on training doppelgängers 320](#)

Table 7.5 [Worksheet for meta-analysis of studies in Table 7.4 322](#)

Table 7.6 [Meta-analysis of hypothetical studies examining the effect of a failure experience on state anxiety 325](#)

Table 7.7 [Meta-analyses performed on the studies of failure and anxiety in Table 7.6 326](#)

## Chapter 9

Table 9.1 [Second-order sampling error: Schematic showing when the two types of second-order sampling error occur 408](#)

## Chapter 12

Table 12.1 [Quantitative results that should be reported in a meta-analysis 492](#)

## Chapter 13

Table 13.1 [Hypothetical example of observed and true mean effect sizes for four sources of effect sizes 519](#)

# List of Figures

## CHAPTER 1

Figure 1.1 [Path model and path coefficients \(From Schmidt, Hunter, & Outerbridge, 1986\)](#) 25

## CHAPTER 2

Figure 2.1 [Path model showing the causal nature of imperfect construct validity in using education as a proxy variable for general cognitive ability](#) 52

Figure 2.2 [Path model showing the assumed causal structure of ability, education, and performance](#) 54

Figure 2.3 [Path model showing the assumed causal structure of ability, income, and education](#) 55

Figure 2.4 [Path model showing the assumed causal structure of teachers' specialized knowledge, teachers' overall GPA, teachers' grade point average in selected education courses, and student performance](#) 57

Figure 2.5 [Path models for cases in which the effect of imperfect construct validity is to attenuate the desired effect size correlation by multiplying the effect size by the construct validity of the proxy variable \( \$r\_{XX'}\$  in both cases\)](#) 60

Figure 2.5a [Extraneous factors in the study independent variable](#) 60

Figure 2.5b [The study as a causal antecedent of the intended independent variable](#) 60

Figure 2.6 [Correctable imperfect construct validity in the dependent variable](#) 61

Figure 2.6a [Extraneous factors in the study measure of the intended dependent variable](#) 61

Figure 2.6b [Using supervisor ratings as a measure of job performance](#) 62

Figure 2.7 [Path diagrams for an extraneous variable produced by the study procedure](#) 65

Figure 2.7a [Path diagram for an extraneous variable Z introduced by the study procedure](#) 65

Figure 2.7b [Experience differences as an extraneous variable produced by the concurrent validation procedure](#) 65

Figure 2.8 [Statistical power: Two examples](#) 73

Figure 2.8a [Statistical power greater than .50](#) 73

Figure 2.8b [Statistical power less than .50](#) 73

## CHAPTER 6

Figure 6.1 [Effect of error of measurement on the separation between the control and experimental groups for a case in which the true score treat effect is  \$\delta = 1.00\$](#)  259

Figure 6.1a [Perfect measurement:  \$r\_{YY} = 1.00\$](#)  259

Figure 6.1b [Typical good measurement:  \$r\_{YY} = .81\$  259](#)

Figure 6.1c [Typical moderate measurement:  \$r\_{YY} = .50\$  259](#)

Figure 6.1d [Typical poor measurement:  \$r\_{YY} = .25\$  259](#)

Figure 6.2 [Path model for the relationship among the delinquency treatment program, the desired measure of actual posttreatment behavior, and the observed posttreatment arrest record 271](#)

Figure 6.3 [Path model for the assumed relationship among the social skills training of supervisors, program mastery, and subsequent interpersonal behavior on the job 272](#)

## CHAPTER 11

Figure 11.1 [Schematic illustrating methods of meta-analysis 465](#)

## CHAPTER 13

Figure 13.1 [Example of funnel plot showing no evidence of availability bias 536](#)

Figure 13.2 [Example of funnel plot showing evidence of availability bias 536](#)

# Preface

---

## Overview of the Third Edition

Meta-analysis is today even more important to the research enterprise than it was when the last edition of this book appeared in 2004 (a fact that is fully explored in [Chapter 1](#)). One reason for this is that meta-analysis is increasingly providing the foundation for evidence-based professional practice in a wide variety of fields, ranging from human resource management to medicine. In recent years, there have been important new methodological developments in meta-analysis, and these are addressed in this book. This book has been prepared to be more user-friendly than the previous editions. Many detailed derivations of equations for which proof no longer need be presented have been eliminated (with references given to sources where they can be found by those interested). In addition, a great deal of effort has been put into making the language of this book as clear as possible to all readers regardless of their statistical and psychometric backgrounds.

Certain new topics appear in multiple chapters throughout the book. For example, there is an expanded treatment of confidence intervals (CIs) in meta-analysis in [Chapters 5](#) (for correlations) and [Chapter 8](#) (for  $d$  values), along with discussion of the difference between the information provided by CIs and credibility intervals (CrIs). And in the example meta-analyses presented in [Chapters 3, 4, 7, and 9](#), CIs are computed for mean observed and corrected effect sizes, along with CrIs. Discussions of the proper interpretation of the percentage or proportion of variance accounted for in meta-analysis are found in these same chapters. The key point is that the square root of the proportion of variance is the correlation between statistical and measurement artifacts and the observed correlations or  $d$  values and that this statistic is more informative than the proportion of variance accounted for.

Many improvements have been made in the package of meta-analysis programs available for applying the methods presented in this book. Some examples include the following. It is now possible to import data files from Excel. All programs now provide CIs around mean observed and corrected effect sizes, as well as CrIs. New features now make it easier to run meta-analyses when there is no range restriction. The upper limit on the number of studies in a meta-analysis has been raised to 1,000. A method of detecting and correcting publication bias (cumulative meta-analysis) has been added. A full list of program improvements can be found in the Appendix to this book. Throughout this book, the example meta-analyses presented are now calculated using the programs in this package.

It is appropriate to point out that the presentations in the book do not rigidly follow statistical conventions. For example,  $\sigma^2$  might appear without the circumflex that indicates that it is not the actual population value 2 but rather an estimate of that value (i.e.,  $\hat{\sigma}^2$ ). We use statistical symbols in this manner for convenience, and in all cases, the context and the text makes it clear which symbols stand for estimates and which stand for actual population parameters. We are aware that our usage of symbols causes some statisticians to tear their hair out.

There have been important improvements and additions in individual chapters. Here we summarize only the major changes. Corrections for measurement error are critical in meta-analysis, but there is often confusion as to the appropriate type of reliability coefficient to use. To address this, [Chapter 3](#) now contains an expanded and more complete treatment of this question to guide users.

[Chapter 4](#) has been greatly simplified by focusing only on the single most accurate method for artifact distribution meta-analysis, the interactive nonlinear (INTNL) method. All example meta-analyses in [Chapter 4](#) are now calculated with the programs for this method. The long technical derivation of the multiplicative method, which has proved to be less accurate, has been dropped. Also, there is now a much clearer presentation of methods for conducting a mixed meta-analysis—one in which one artifact is corrected individually in each study while the other artifacts are corrected based on artifact distributions. This presentation is accompanied by a detailed worked numerical example.

In [Chapter 5](#), the long technical derivation of the equations for correcting indirect range restriction has been dropped, because that material has since been published in a major journal (Hunter, Schmidt, & Le, 2006). [Chapter 5](#) now presents an expanded discussion of the issues involved in path analyses based on meta-analytic correlation matrices, an application that has now become common in the literature. There is also an updated and expanded discussion of the problems created by use of Fisher's  $z$  in meta-analysis of correlations. In [Chapter 6](#), only minimal changes have been necessary.

[Chapter 7](#) now contains an improved presentation of methods for conducting meta-analysis of  $d$  values when artifacts beyond sampling error and measurement error in the dependent variable measure are corrected for. There is a clear explanation of how to convert  $d$  values to point-biserial correlations, conduct the meta-analysis in the  $r$  statistic metric, and then convert the meta-analysis results back into the  $d$  statistic metric. This approach to meta-analysis of  $d$  values has become increasingly common in the literature.

Important changes have been made in [Chapter 8](#). The most common experimental design, the independent

groups design, is covered in detail in [Chapter 7](#). But many other experimental designs are encountered in the literature. [Chapter 8](#) now presents the needed equations for computing the appropriate  $d$  values from these designs and the equations for the sampling error variance of the resulting  $d$  values. These methods allow  $d$  values from these experimental designs to be included in the meta-analysis along with the  $d$  values from the independent groups design. An equation is also presented for adjusting the sample sizes of the different designs, so that the  $d$  value and the adjusted  $N$  can be entered into the meta-analysis programs written for the independent groups design studies. This step ensures that the sampling error variances computed by the programs are correct.

[Chapter 9](#) (“General Technical Issues in Meta-Analysis”) contains much new material. This chapter examines in detail the issues involved in detecting and calibrating moderators in meta-analysis. This includes discussion of subgrouping of studies, hierarchical meta-analysis via subgrouping, multilevel and hierarchical linear models (HLM) for meta-analysis, mixed models in meta-analysis, and meta-regression. Statistical problems in the use of meta-regression, which are often ignored in published applications, are discussed in detail. There is an in-depth discussion of the debate over optimal study weights in meta-analysis. This chapter also presents a detailed discussion of second-order sampling error in meta-analysis and a new method for conducting second-order meta-analysis, along with two exercises applying this new method. (Other meta-analysis exercises are provided at the end of [Chapters 3](#), [4](#), and [7](#).) In connection with confidence intervals, there is an explanation of the differences in how these are computed in the Hunter-Schmidt and the Hedges-Vevea meta-analysis methods. Finally, there is a discussion of the use of the odds ratio statistic in psychology and social science research.

[Chapter 11](#) presents an updated discussion of different methods of meta-analysis, including one recently developed approach (SEM-based meta-analysis). There is also updated information on software available for conducting meta-analyses using different methods.

[Chapter 12](#) includes a new discussion of how the results of meta-analyses have been and should be presented in research reports and published articles. Reporting standards in meta-analysis are important, and the evidence indicates that they need improvement.

[Chapter 13](#) presents a greatly expanded treatment of source bias and publication bias, an area that has recently received much attention in the literature. Because of the potential effects on the accuracy of meta-analyses in the literature, the recent research and findings on research fraud and QRPs (questionable research practices), which result in biased results in primary studies, are also explored. Next, there is a

discussion and evaluation of nine different methods of detecting (and sometimes correcting) publication and source bias. Finally, there is a discussion of software for detecting publication bias.

The methods presented in this book differ from other approaches to meta-analysis in a wide variety of technical ways. However, all of these differences stem from one central difference: a difference in how the purpose of meta-analysis is defined. The avowed purpose of other methods of meta-analysis is to describe and summarize the results reported by studies in a given research literature (Rubin, 1990). The purpose of our methods is very different. In our view, the purpose of meta-analysis is to estimate what the results would have been if all the studies had been conducted without methodological limitation or flaws. The results of perfectly conducted studies would reveal the underlying construct-level relationships—that is, the true “state of nature.” It is these relationships that we as scientists are interested in, as Rubin (1990) and others have noted. Our methods estimate these relationships. We are much less interested in attaining an accurate description and summary of the results reported by necessarily flawed primary studies. This critical distinction is developed in more detail near the end of [Chapter 1](#) and in [Chapter 14](#).

---

## Brief History of this Book

This book is labeled as the third edition. In actuality, it should be considered the fourth edition. The first edition was really the Hunter, Schmidt, and Jackson (1982) book. This was followed by Hunter and Schmidt (1990b) and Hunter and Schmidt (2004), and now Schmidt and Hunter (2014). Gene V. Glass published the first journal article on meta-analysis in 1976 in an education journal (*The Educational Researcher*). In that article, he laid out the essential rationale and defined many of the basic features of meta-analysis as it is known today. He also coined the label *meta-analysis*. Our work was not in educational psychology but in industrial and organizational psychology. Unaware of Glass's work, we developed our meta-analysis methods in 1975 and applied them to empirical data sets from personnel selection research. But instead of immediately submitting our report for publication, we submitted it to the James McKeen Cattell Research Design Contest sponsored by Division 14 (The Society for Industrial and Organizational Psychology) of the American Psychological Association. To be eligible for this award, entries must not have already been published or accepted for publication (i.e., in press). Our development and initial application of meta-analysis (then called “validity generalization”) won the Cattell award for 1976, but the 1-year delay in publication (Schmidt & Hunter, 1977) meant that our first meta-analysis article appeared 1 year after Glass's. Glass's

1976 article was not only the first in time but also the first to emphasize meta-analysis as a fully general set of methods that should be applied to the integration of research literatures in all areas. At that time, our major emphasis was on solving the problem of the apparent substantial variability of test validities in the personnel selection literature (see [Chapter 4](#)). But we were aware of the potential of our methods for application to other research literatures. So when Lee J. Cronbach suggested to us in correspondence in early 1978 that our methods could be applied to research literatures in many areas of the behavioral and social sciences, we had already begun to think about a possible book presenting our methods in that light. That book was published in 1982 (Hunter et al., 1982). (But again Glass was first; he and his coauthors published their meta-analysis book in 1981; Glass, McGaw, & Smith, 1981.) Since then, the methods initially presented in our 1982 book have been widely applied. As indicated earlier, that book was followed by three subsequent editions, including this book.

An edited book devoted entirely to the history and impact of our meta-analysis methods is available (Murphy, 2003). This book is a good source of information beyond that contained in this book. In particular, [Chapter 2](#) of that book, by the authors of this book, presents a history of the development of these methods. A history of these methods will also be published in the journal *Research Synthesis Methods* in 2014 (Schmidt, in press). DeGeest and Schmidt (2011) present a detailed account of the impact of these methods on cumulative knowledge in the areas of industrial-organizational psychology, human resources, management, and organizational behavior.

---

## Organization of this Book

In scientific reports, a review of earlier developments is usually presented first. In the case of this book, that would be a review of previous methods of integrating research literatures (pre-meta-analysis methods). But in the case of this topic, such a review would be difficult for readers to fully understand without first having knowledge of meta-analysis principles and methods. Therefore, we first present the methods of psychometric meta-analysis in some detail. Then later in the book ([Chapter 11](#)), we present a review and critique of other research integration methods.

In the course of conducting a meta-analysis on a particular relationship, the temporal sequence of steps is as follows: (1) search for and gather studies, (2) extract and code information from the studies, (3) apply meta-analysis methods to the information extracted, and (4) present the results in a report or article. This



book discusses all four steps, but not in their natural chronological order. The reason for this is that in order to know what is required in the first two steps, you must know in detail what you are going to do in the third step. Therefore, we cover meta-analysis methods first and then return to issues of defining the study domain, locating studies, deciding what to code, and writing up the meta-analysis report. Finally, we go back one step further and present recommendations for improvements in reporting practices in primary studies that are necessary for optimal application of meta-analysis methods to those studies.

My coauthor, John (Jack) Hunter, died on June 26, 2002. His contributions to the development of the methods presented in this book were beyond enumeration. The story of his life is sketched in Schmidt (2003). To me, he was not only an inspired collaborator of 30 years but also a best friend. The great loss to psychology and the other social sciences is in addition to my personal loss. This book is dedicated to his memory. Of course, all errors and omissions in this edition (and in the 2004 book; the second edition) are mine. I would like to believe that he would approve of this third edition.

*Frank L. Schmidt*

## Acknowledgments

Jack Hunter has now been gone 12 years, but his contributions to the psychometric meta-analysis methods live on in the third edition of this book, and I want to give them first place in any acknowledgments. I would like to thank all of my colleagues and PhD students whose persistent curiosity and questions about meta-analysis stimulated the development of many ideas in this book. I would also like to thank them for urging me to revise and update the 2004 edition of this book and for encouraging me to finish it once I had started. The wait has been long, but I hope they will be happy with the final product. Special thanks go to In-Sue Oh, Huy Le, Michael McDaniel, Deniz Ones, Hannah Rothstein, Vish Viswesvaran, and Kenneth S. Law for their insightful comments on draft chapters. I would like to thank Vicki Knight, who arranged the contract for this revision and assisted along the way, as well as our production editor, Laura Barrett, and our copy editor, Gillian Dickens, for their patience, professionalism, and support.

Thank you to all of the reviewers for their time and input: Christopher M. Berry, Texas A&M University; Terri D. Pigott, Loyola University Chicago; Steven Pulos, University of Northern Colorado; Emily E. Tanner-Smith, Vanderbilt University; and Jeffrey B. Vancouver, Ohio University.

I owe a special debt to my wife, Cindy, for her support and encouragement during the long process of preparing this new edition. Finally, we would like to thank Linda Bostian for her dedication, skill, and professionalism in preparing this manuscript, even when the work extended to weekends and afterhours.

## About the Authors

**Frank L. Schmidt** is the Gary F. Fethke Leadership Professor Emeritus in the Department of Management and Organization in the Tippie College of Business at the University of Iowa. He received his PhD in industrial-organizational psychology from Purdue University and has been on the faculties of Michigan State and George Washington Universities. He has authored or coauthored seven books and nearly 200 articles and book chapters on measurement, statistics, research methods, individual differences, and personnel selection. He headed a research program in the U.S. Office of Personnel Management in Washington, D.C., for 11 years, during which time he published numerous research studies in personnel psychology, primarily with John Hunter. Their research on the generalizability of employment selection method validities led to the development of the meta-analysis methods presented in this book. Professor Schmidt has received the Distinguished Scientific Award for Contributions to Applied Psychology from the American Psychological Association (APA) (jointly with John Hunter) and the Distinguished Scientific Contributions Award from the Society for Industrial and Organizational Psychology (SIOP) (also jointly with John Hunter). He has also received the Ingram Olkin Award and the Frederick Mosteller Award, both for contributions to meta-analysis methodology; the Scientific Award for Applications of Psychology from the Association for Psychological Science; the Gold Medal Lifetime Achievement award from the APA Foundation; and the Distinguished Career Award for Contributions to Human Resources and the Distinguished Career Achievement Award for Contributions to Research Methods, both from the Academy of Management. He is a Fellow of the APA, the Association for Psychological Science, and SIOP and is past president of Division 5 (Measurement, Statistics, & Evaluation) of the APA.

**John E. (Jack) Hunter** (1939–2002) was a professor in the Department of Psychology at Michigan State University. He received his PhD in quantitative psychology from the University of Illinois. Jack coauthored four books and authored or coauthored more than 200 articles and book chapters on a wide variety of methodological topics, including confirmatory and exploratory factor analysis, measurement theory and methods, statistics, and research methods. He also published numerous research articles on such substantive topics as intelligence, attitude change, the relationship between attitudes and behavior, validity generalization, differential validity/selection fairness, and selection utility. Much of his research on attitudes was in the field of communications, and the American Communications Association named a research award in his honor. Professor Hunter received the Distinguished Scientific Award for Contributions to Applied Psychology from the American Psychological Association (APA) (jointly with Frank Schmidt) and the Distinguished Scientific Contributions Award from the Society for Industrial and Organizational Psychology

(SIOP) (also jointly with Frank Schmidt). He was a Fellow of APA, Association for Psychological Science, and SIOP and was a past president of the Midwestern Society for Multivariate Experimental Psychology. For the story of Jack's life, see Schmidt (2003).

# Introduction to Meta-Analysis

1. Integrating Research Findings Across Studies	3
2. Study Artifacts and Their Impact on Study Outcomes	39

# Integrating Research Findings across Studies

Before we delve into a discussion of methods, we would like to consider a concrete example. The next section presents a set of studies to be reviewed, then a sample narrative review, followed by a critique of this review. It has been our experience that personal experience with the problems of such a review greatly enhances the learning process.

## General Problem and an Example

A major task in all areas of science is the development of theory. In many cases, the theorists have available the results of a number of previous studies on the subject of interest. Their first task is to find out what empirical relationships have been revealed in these studies so they can take them into account in theory construction. In developing an understanding of these relationships, it is often helpful in reviewing the studies to make up a table summarizing the findings of these studies. [Table 1.1](#) shows such a summary table put together by a psychologist attempting to develop a theory of the relationship between job satisfaction and organizational commitment. In addition to the observed correlations and their sample sizes, the psychologist has recorded data on (1) sex, (2) organization size, (3) job level, (4) race, (5) age, and (6) geographical location. The researcher believes variables 1, 2, 3, and 4 may affect the extent to which job satisfaction gets translated into organizational commitment. The researcher has no hypotheses about variables 5 and 6 but has recorded them because they were often available.

Table 1.1 Correlations between Organizational Commitment and Job Satisfaction.								
Study	N	r	Sex	Size of Organization	White vs. Blue Collar	Race	Under vs. Over 30	North Vs. South
1	20	.46*	F	S	WC	B	U	N
2	72	.32**	M	L	BC	Mixed	Mixed	N
3	29	.10	M	L	WC	W	O	N
4	30	.45**	M	L	WC	W	Mixed	N

5	71	.18	F	L	BC	W	O	N
6	62	.45**	F	S	BC	W	U	N
7	25	.56**	M	S	BC	Mixed	U	S
8	46	.41**	F	L	WC	W	Mixed	S
9	22	.55**	F	S	WC	B	U	N
10	69	.44**	F	S	BC	W	U	N
11	67	.34**	M	L	BC	W	Mixed	N
12	58	.33**	M	S	BC	W	U	N
13	23	.14	M	S	WC	B	O	S
14	20	.36	M	S	WC	W	Mixed	N
15	28	.54**	F	L	WC	W	Mixed	S
16	30	.22	M	S	BC	W	Mixed	S
17	69	.31**	F	L	BC	W	Mixed	N
18	59	.43**	F	L	BC	W	Mixed	N
19	19	.52*	M	S	BC	W	Mixed	S
20	44	-.10	M	S	WC	W	O	N
21	60	.44**	F	L	BC	Mixed	Mixed	N
22	23	.50**	F	S	WC	W	Mixed	S
23	19	-.02	M	S	WC	B	O	S
24	55	.32**	M	L	WC	W	Mixed	Unknown

25	19	.19	F	S	WC	B	O	N
26	26	.53**	F	S	BC	B	U	S
27	58	.30*	M	L	WC	W	Mixed	S
28	25	.26	M	S	WC	W	U	S
29	28	.09	F	S	BC	W	O	N
30	26	.31	F	S	WC	Mixed	U	S

\*  $p < .05$ .

\*\*  $p < .01$ .

As an exercise in integrating findings across studies and constructing theory, we would like you to spend a few minutes examining and interpreting the data in [Table 1.1](#). We would like you to jot down the following:

1. The tentative conclusions you reached about the relationship between job satisfaction and organizational commitment and the variables that do and do not moderate that relationship
2. An outline of your resulting theory of this relationship

### A Typical Interpretation of the Example Data

A typical report on the findings shown in [Table 1.1](#) would run like this: The correlation between occupational commitment and job satisfaction varies from study to study with the correlation varying between  $-.10$  and  $.56$ . Although 19 out of 30 studies found a significant correlation, 11 of 30 studies found no relationship between commitment and satisfaction.

For male work populations, commitment and satisfaction were correlated in 8 studies and not correlated in 7 (i.e., correlated in 53% of the studies), while for women there was a correlation in 11 of 15 cases (or in 73% of the studies). Correlation was found in 83% of the large organizations but in only 50% of the small organizations. Correlation was found in 79% of the blue-collar populations but in only 50% of the white-collar populations. Correlation was found in 67% of the populations that were all white or mixed race, while correlation was found in only 50% of those work populations that were all black. Correlation was found in 83% of the



cases in which the workforce was all younger than 30 or a mixture of younger and older workers, while not a single study with only older workers found a significant correlation. Finally, 65% of the studies done in the north found a correlation, while only 58% of the southern studies found a correlation. Each of the differences between work populations could be taken as the basis for a hypothesis that there is an interaction between that characteristic and organizational commitment in the determination of job satisfaction.

If the studies done on older workers are removed, then significant correlation is found for 19 of the remaining 23 studies. Within these 23 studies with younger or mixed-age work populations, all 10 correlations for large organizations were significant.

There are 13 studies of younger or mixed-age work populations in small organizations. Within this group of studies, there is a tendency for correlation between organizational commitment and job satisfaction to be more likely found among women, among blue-collar workers, in all-black work populations, and in the north.

### **Conclusions of the Review**

Organizational commitment and job satisfaction are correlated in some organizational settings but not in others. In work groups in which all workers are older than 30, the correlation between commitment and satisfaction was never significant. For young or mixed-age work populations, commitment and satisfaction are always correlated in large organizations. For young or mixed-age work populations in small organizations, correlation was found in 9 of 13 studies with no organizational feature capable of perfectly accounting for those cases in which correlation was not found.

These findings are consistent with a model that assumes that organizational commitment grows over about a 10-year period to a maximum value at which it asymptotes. Among older workers, organizational commitment may be so uniformly high that there is no variation. Hence, among older workers, there can be no correlation between commitment and job satisfaction. The finding for large organizations suggests that growth of commitment is slower there, thus generating a greater variance among workers of different ages within the younger group.

### Critique of the Sample Review

The preceding review was conducted using review practices that characterize many narrative review articles not only in psychology but in sociology, education, and the rest of the social sciences as well. Yet every conclusion in the review is false. The data were constructed by a Monte Carlo run in which the population correlation was always .33. After a sample size was randomly chosen from a distribution centering about 40, an observed correlation was chosen using the standard distribution for  $r$  with mean  $\rho = .33$  and variance

$$\frac{(1 - \rho^2)^2}{N - 1}.$$

That is, the variation in results in [Table 1.1](#) is entirely the result of sampling error. Each study is conducted on a small sample and hence generates an observed correlation that departs by some random amount from the population value of .33. The size of the departure depends on the sample size. Note that the largest and smallest values found in [Table 1.1](#) are all from studies with very small samples. The larger sample size studies tend to show less of a random departure from .33.

The moderator effects appear to make sense, yet they are purely the results of chance (i.e., sampling error). The values for the organizational characteristics were assigned to the studies randomly.

The crucial lesson to be learned from this exercise is this: “Conflicting results in the literature” may be entirely artifactual. The data in [Table 1.1](#) were generated by using one artifact for generating false variation across studies, sampling error. There are other artifacts that are found in most sets of studies: Studies vary in terms of the quality of measurement (reliability) of their scales; researchers make computational or computer errors; people make typographical errors in copying numbers from computer output or in copying numbers from handwritten tables onto manuscripts or in setting tables into print; researchers study variables in settings with greater or smaller ranges of individual differences (range variation); and so on. In our experience (described later), many of the interactions hypothesized to account for differences in findings in different studies are nonexistent; that is, they are apparitions composed of the ectoplasm of sampling error and other artifacts.

## Problems with Statistical Significance Tests

In the data set given in [Table 1.1](#), all study population correlations are actually equal to .33. Of the 30 correlations, 19 were found to be statistically significant. However, 11 of the 30 correlations were not significant. That is, the significance test gave the wrong answer 11 out of 30 times, an error rate of 37%. Many people express shock that the error rate can be greater than 5%. The significance test was derived in response to the problem of sampling error, and many believe that the use of significance tests guarantees an error rate of 5% or less. This belief is false. Statisticians have pointed this out for many years; the possibility of high error rates is brought out in discussions of the “power” of statistical tests. However, statistics instructors are well aware that this point is not understood by most students. The 5% error rate is guaranteed only if the null hypothesis is true. If the null hypothesis is false, then the error rate can go as high as 95%.

Let us state this in more formal language. If the null hypothesis is true for the population and our sample data lead us to reject it, then we have made a Type I error. If the null hypothesis is false for the population and our sample data lead us to accept it, then we have made a Type II error. The statistical significance test is defined in such a way that the Type I error rate is at most 5%. However, the Type II error rate is typically left free to be as high as 95%. The question is which error rate applies to a given study. The answer is that the relevant error rate can only be known if we know whether the null hypothesis is true or false for that study. If we know that the null hypothesis is true, then we know that the significance test has an error rate of 5%. Of course, if we know that the null hypothesis is true and we still do a significance test, then we should wear a dunce cap, because if we know the null hypothesis to be true, then we can obtain a 0% error rate by ignoring the data. That is, there is a fundamental circularity to the significance test. If you do not know whether the null hypothesis is true or false, then you do not know whether the relevant error rate is Type I or Type II; that is, you do not know if your error rate is 5% or some value as high as 95%. There is only one way to guarantee a 5% error rate in all cases: Abandon the significance test and use a confidence interval.

Consider our hypothetical example from [Table 1.1](#). However, let us simplify the example still further by assuming that the sample size is the same for all studies, say  $N = 40$ . The one-tailed significance test for a correlation coefficient is  $\sqrt{N-1}r \geq 1.64$ ; in our case,  $\sqrt{39}r \geq 1.64$  or  $r \geq .26$ . If the population correlation is .33 and the sample size is 40, the mean of the sample correlations is .33, while the standard deviation is  $(1 - \rho^2) / \sqrt{N-1} = (1 - .33^2) / \sqrt{39} = .14$ . Thus, the probability that the observed correlation will be significant is the probability that the sample correlation will be greater than .26 when the population value is .33 and a

standard deviation of .14:

$$P\{r \geq .26\} = P\left\{\frac{r - .33}{.14} \geq \frac{.26 - .33}{.14}\right\} = P\{z \geq -.50\} = .69.$$

That is, if all studies were done with a sample size of 40, then a population correlation of .33 would mean an error rate of 31% (i.e.,  $1 - .69 = .31$ ).

Suppose we alter the population correlation in our hypothetical example from .33 to .20. Then the probability that the observed correlation will be significant drops from .69 to

$$P\{r \geq .26\} = P\left\{z \geq \frac{.26 - .20}{.15} = .39\right\} = .35.$$

That is, the error rate rises from 31% to 65%. In this realistic example, we see that the error rate can be over 50%. A two-to-one majority of the studies can find the correlation to be not significant despite the fact that the population correlation is always .20.

Error rates of 50% or higher have been shown to be the usual case in many research literatures. Thus, reviewers who count the number of significant findings are prone to incorrectly conclude that a relationship does not exist when it does. Furthermore, as Hedges and Olkin (1980) pointed out, this situation only gets worse as more studies are conducted. The reviewer will become ever more convinced that the majority of studies show no effect and that the effect thus does not exist. Statistical power has been examined in much of the research literature in psychology, starting with Cohen (1962) and extending up to the present. In most literatures, the mean statistical power is in the .40 to .60 range and is as low as .20 in some areas (Hunter, 1997; Schmidt, 1996; Schmidt & Hunter, 2003; Sedlmeier & Gigerenzer, 1989). And, surprisingly, over time there has been little or no increase in statistical power in published studies (Sedlmeier & Gigerenzer, 1989. Maxwell (2004) explores the reasons why this is the case. Other possible reasons are given in [Chapter 13](#).

If the null hypothesis is true in a set of studies, then the base rate for significance is not 50% but 5%. If more than 1 in 20 studies finds significance, then the null hypothesis must be false in some studies. We must then avoid an error made by some reviewers who know the 5% base rate. For example, if 35% of the findings are significant, some have concluded that "Because 5% will be significant by chance, this means that the number of studies in which the null hypothesis is truly false is  $35 - 5 = 30\%$ ." Our hypothetical example shows this reasoning to be false. If the population correlation is .20 in every study and the sample size is always 40, then there will be significant findings in only 35% of the studies, even though the null hypothesis is false in all cases.

The typical use of significance test results leads to gross errors in traditional review studies. Most such reviews falsely conclude that further research, focused on moderator variables, is needed to resolve the “conflicting results” in the literature. These errors in review studies can only be eliminated if errors in the interpretation of significance tests can be eliminated. Yet those of us who have been teaching power to generation after generation of graduate students have been unable to change the reasoning processes and the false belief in the 5% error rate (Sedlmeier & Gigerenzer, 1989).

This example illustrates a critical point. Traditional reliance on statistical significance tests in interpreting studies leads to false conclusions about what the study results mean; in fact, the traditional approach to data analysis makes it virtually impossible to reach correct conclusions in most research areas (Hunter, 1997; Schmidt, 1996, 2010).

A common reaction to the preceding critique of traditional reliance on significance testing goes something like this: “Your explanation is clear, but I don’t understand how so many researchers (and even some methodologists) could have been so wrong so long on a matter as important as the correct way to analyze data? How could psychologists and other researchers have failed to see the pitfalls of significance testing?” Over the years, a number of methodologists have addressed this question (Carver, 1978; Cohen, 1994; Guttman, 1985; Meehl, 1978; Oakes, 1986; Rozeboom, 1960; Schmidt & Hunter, 1997). In their statistics classes, young researchers have typically been taught a lot about Type I error and very little about Type II error and statistical power. Thus, they are unaware that the error rate is very large in the typical study; they tend to believe the error rate is the alpha level used (typically .05 or .01). In addition, empirical research suggests that most researchers believe that the use of significance tests provides them with many nonexistent benefits in understanding their data. For example, most researchers believe that a statistically significant finding is a “reliable” finding in the sense that it will replicate if a new study is conducted (Carver, 1978; Oakes, 1986; Schmidt, 1996; Schmidt & Hunter, 1997). For example, they believe that if a result is significant at the .05 level, then the probability of replication in subsequent studies (if conducted) is  $1.00 - .05 = .95$ . This belief is completely false. The probability of replication is the statistical power of the study and is almost invariably much lower than .95 (e.g., typically .50 or less). Killeen (2005a, 2005b) proposed a statistic called *P-rep* that he claimed gives the probability that a research finding would be replicated in a new study. For a few years, this statistic was widely used. However, Trafimow, MacDonald, Rice, and Carlson (2010) demonstrated mathematically that the *P-rep* statistic did not in fact provide this probability. Today *P-rep* is rarely used.

Most researchers also believe that if a result is nonsignificant, one can conclude that it is probably just due

to chance, another false belief, as illustrated in our example in which all nonsignificant results were Type II errors. There are other widespread but false beliefs about the usefulness of information provided by significance tests (Carver, 1978; Oakes, 1986). Discussion of these beliefs can be found in Schmidt (1996) and Schmidt and Hunter (1997).

Another fact is relevant at this point: The physical sciences, such as physics and chemistry, do not use statistical significance testing in interpreting their data (Cohen, 1990). Instead, they use confidence intervals. It is no accident, then, that these sciences have not experienced the debilitating problems described here that are inevitable when researchers rely on significance tests. Given that the physical scientists regard reliance on significance testing as unscientific, it is ironic that so many psychologists defend the use of significance tests on grounds that such tests are the objective and scientifically correct approach to data analysis and interpretation. In fact, it has been our experience that psychologists and other behavioral scientists who attempt to defend significance testing usually equate null hypothesis statistical significance testing with scientific hypothesis testing in general. They argue that hypothesis testing is central to science and that the abandonment of significance testing would amount to an attempt to have a science without hypothesis testing. They falsely believe that significance testing and hypothesis testing in science are one and the same thing. This belief is tantamount to stating that physics, chemistry, and the other physical sciences are not legitimate sciences because they do not test their hypotheses using statistical significance testing. Another logical implication of this belief is that prior to the introduction of null hypothesis significance testing by R. A. Fisher (1932) in the 1930s, no legitimate scientific research was possible. The fact is, of course, that there are many ways to test scientific hypotheses—and that significance testing is one of the least effective methods of doing this (Schmidt & Hunter, 1997).

---

## Is Statistical Power the Solution?

Some researchers believe that the only problem with significance testing is low power and that if this problem could be solved there would be no problems with reliance on significance testing. These individuals see the solution as larger sample sizes. They believe that the problem would be solved if every researcher, before conducting each study, would calculate the number of subjects needed for “adequate” power (usually taken as power of .80) and then would use that sample size. What this position overlooks is that this requirement would make it impossible for most studies ever to be conducted. At the start of research in a given area, the

questions are often of the form “Does Treatment A have an effect?” (e.g., Does interpersonal skills training have an effect? Does cognitive behavior therapy work?). If Treatment A indeed has a substantial effect, the sample size needed for adequate power may not be prohibitively large. But as research develops, subsequent questions tend to take the form “Is the effect of Treatment A larger than the effect of Treatment B?” (e.g., Is the effect of the new method of training larger than that of the old method? Is Predictor A more valid than Predictor B?). The effect size then becomes the *difference* between the two effects. Such effect sizes will often be small, and the required sample sizes are therefore often quite large—1,000 or 2,000 or more (Schmidt & Hunter, 1978). And this is just to attain power of .80, which still allows a 20% Type II error rate when the null hypothesis is false—an error rate most would consider high. Many researchers cannot obtain that many subjects, no matter how hard they try; either it is beyond their resources or the subjects are just not available at any cost. Thus, the upshot of this position would be that many—perhaps most—studies would not be conducted at all.

People advocating the power position say this would not be a loss. They argue that a study with inadequate power cannot support a research conclusion and therefore should not be conducted. Such studies, however, contain valuable information when combined with others like them in a meta-analysis. In fact, accurate meta-analysis results can be obtained based on studies that *all* have inadequate statistical power individually, because meta-analysis can provide precise estimates of average effect size. The information in these studies is lost if these studies are never conducted.

The belief that such studies are worthless is based on two false assumptions: (1) the assumption that every individual study must be able to justify a conclusion on its own, without reference to other studies, and (2) the assumption that every study should be analyzed using significance tests. One of the contributions of meta-analysis has been to show that no single study is adequate by itself to answer a scientific question. Therefore, each study should be considered as a data point to be contributed to a later meta-analysis. In addition, individual studies should be analyzed using not significance tests but point estimates of effect sizes and confidence intervals.

How, then, *can* we solve the problem of statistical power in individual studies? Actually, this problem is a pseudoproblem. It can be “solved” by discontinuing the significance test. As Oakes (1986, p. 68) noted, statistical power is a legitimate concept only within the context of statistical significance testing. If significance testing is not used, then the concept of statistical power has no place and is not meaningful. In particular, there need be no concern with statistical power when point estimates and confidence intervals are used to

analyze data in studies and meta-analysis is used to integrate findings across studies.

Our critique of the traditional practice of reliance on significance testing in analyzing data in individual studies and in interpreting research literature might suggest a false conclusion, namely, that if significance tests had never been used, the research findings would have been consistent across different studies examining a given relationship. Consider the correlation between job satisfaction and job performance in [Table 1.1](#). Would these studies have all had the same findings if researchers had not relied on significance tests? Absolutely not: The correlations would have varied widely (as indeed they did). The major reason for such variability in correlations is simple sampling error—caused by the fact that the small samples used in individual research studies are randomly unrepresentative of the populations from which they are drawn. Most researchers severely underestimate the amount of variability in findings that is caused by sampling error.

The law of large numbers correctly states that large random samples are representative of their populations and yield parameter estimates that are close to the actual population values. Many researchers seem to believe that the same law applies to small samples. As a result, they erroneously expect statistics computed on small samples (e.g., 50 to 300) to be close approximations to the real (population) values. In one study we conducted (Schmidt, Ocasio, Hillery, & Hunter, 1985), we drew random samples (small studies) of  $N = 30$  from a much larger single data set ( $N = 1,455$ ;  $r = .22$ ) and computed results on each  $N = 30$  sample. The resulting validity estimates varied dramatically from “study” to “study,” ranging from  $-.21$  to  $.61$ , with all this variability being due solely to sampling error (Schmidt, Ocasio, et al., 1985). Yet when we showed these data to researchers, they found it hard to believe that each “study” was a random draw from the same larger study. They did not believe simple sampling error could produce that much variation. They were shocked because they did not realize how much variation simple sampling error produces in research studies.

There are two alternatives to the significance test. At the level of review studies, there is meta-analysis. At the level of single studies, there is the confidence interval.

---

## Confidence Intervals

Consider Studies 17 and 30 from our hypothetical example in [Table 1.1](#). Study 17, with  $r = .31$  and  $N = 69$ , finds the correlation to be significant at the .01 level. Study 30, with  $r = .31$  and  $N = 26$ , finds the correlation to be not significant. That is, two authors with an identical finding,  $r = .31$ , come to opposite conclusions. Author



17 concludes that organizational commitment is highly related to job satisfaction, while Author 30 concludes that they are unrelated. Thus, two studies with identical findings can lead to a review author claiming “conflicting results in the literature.”

The conclusions are quite different if the results are interpreted with confidence intervals. Author 17 reports a finding of  $r = .31$  with a 95% confidence interval of  $.10 \leq \rho \leq .52$ . Author 30 reports a finding of  $r = .31$  with a 95% confidence interval of  $-.04 \leq \rho \leq .66$ . There is no conflict between these results; the two confidence intervals overlap substantially.

Consider now Studies 26 and 30 from [Table 1.1](#). Study 26 finds  $r = .53$  with  $N = 26$ , which is significant at the .01 level. Study 30 finds  $r = .31$  with  $N = 26$ , which is not significant. That is, we have two studies with the same sample size but apparently widely divergent results. Using significance tests, one would conclude that there must be some moderator that accounts for the difference. This conclusion is false.

Had the two studies used confidence intervals, the conclusion would have been different. The confidence interval for Study 26 is  $.25 \leq \rho \leq .81$ , and the confidence interval for Study 30 is  $-.04 \leq \rho \leq .66$ . It is true that the confidence interval for Study 30 includes  $\rho = 0$ , while the confidence interval for Study 26 does not; this is the fact registered by the significance test. The crucial thing, however, is that the two confidence intervals show an overlap of  $.25 \leq \rho \leq .66$ . Thus, consideration of the two studies together leads to the correct conclusion that it is possible that both studies could imply the same value for the population correlation  $\rho$ . Indeed, the overlapping intervals include the correct value,  $\rho = .33$ .

Two studies with the same population value can have non-overlapping confidence intervals, but this is a low-probability event (about 5%). But, then, confidence intervals are not the optimal method for looking at results across studies; that distinction belongs to meta-analysis.

Confidence intervals are more informative than significance tests for two reasons. First, the interval is correctly centered on the observed value rather than on the hypothetical zero value of the null hypothesis. Second, the confidence interval gives the researcher a correct picture of the extent of uncertainty in small sample studies. It may be disconcerting to see a confidence interval as wide as  $-.04 \leq \rho \leq .66$ , but that is far superior to the frustration produced over the years by the false belief in “conflicting results.”

Confidence intervals can be used to generate definitions for the phrase “small sample size.” Suppose we want the confidence interval for the correlation coefficient to define the correlation to the first digit, that is, to have a width of  $\pm .05$ . Then, for small population correlations, the minimum sample size is approximately 1,538. For

a sample size of 1,000 to be sufficient, the population correlation must be at least .44. Thus, under this standard of accuracy, for correlational studies, “small sample size” includes all studies with less than a thousand persons and often extends above that.

There is a similar calculation for experimental studies. If the statistic used is the  $d$  statistic (by far the most frequent choice), then small effect sizes will be specified to their first digit only if the sample size is 3,076. If the effect size is larger, then the sample size must be even greater than 3,076. For example, if the difference between the population means is .30 standard deviations, then the minimum sample size that yields accuracy to within  $\pm .05$  of .30 is 6,216. Thus, given this standard of accuracy, for experimental studies, “small sample size” begins with 3,000 and often extends well beyond that. Now think about the fact that many, perhaps most, experimental studies in behavior labs have total  $N$ s between 20 and 50.

Since the publication of the first edition of this book in 1990, recognition of the superiority of confidence intervals and point estimates of effect sizes over significance tests has grown exponentially. The report of the task force on significance testing of the American Psychological Association (APA) (Wilkinson & The APA Task Force on Statistical Inference, 1999) stated that researchers should report effect size estimates and confidence intervals. The fifth and sixth editions of the *APA Publication Manual* stated that it is almost always necessary for primary studies to report effect size estimates and confidence intervals (American Psychological Association, 2001, 2009). Twenty-one research journals in psychology and education now require that these statistics be reported (B. Thompson, 2002). Some have argued that information on the methods needed to compute confidence intervals is not widely available. However, there are now helpful and informative statistics textbooks designed around point estimates of effect size and confidence intervals instead of significance testing (Cumming, 2012; Kline, 2004; Lockhart, 1998; Smithson, 2000). The Cumming (2012) book includes excellent online computer programs that make calculations easy and that illustrate critical statistical facts and principles. B. Thompson (2002) presents considerable information on computation of confidence intervals and cites many useful references that provide more detail (e.g., Kirk, 2001; Smithson, 2001). The August 2001 issue of *Educational and Psychological Measurement* was devoted entirely to methods of computing and interpreting confidence intervals. There are many other such publications (e.g., Borenstein, 1994).

Despite these developments, most published articles still use significance tests. How this can be is something of a mystery, given the fact that this practice has been completely discredited. Orlitzky (2011) argues that the problem is that the evidence against significance testing has not been *institutionalized*. Articles discrediting significance testing have been aimed at inducing individual researchers to change their statistical practices,

not at a broader, more systematic or institutional change. But it very difficult for individual researchers to go against what has become an institutionalized practice in most journals. What is needed, he contends, is top-down disciplinary-wide changes in the research culture. This is a broad recommendation. For example, he says that urging individual journal editors to require effect sizes and confidence intervals will produce little change. There must be an enforceable agreement at the level of the entire discipline that proper data analysis procedures must be used in primary studies, along with major changes in the way research methods are taught in graduate programs. This is a long-term proposition for culture change. Fortunately as we will see next, meta-analysis makes it possible to make progress in developing cumulative knowledge in the interim even if significance testing continues to be used in individual primary studies.

---

## Meta-Analysis

Is there a quantitative analysis that would have shown that all the differences in [Table 1.1](#) might stem from sampling error? Suppose we compute the variance of the correlations, weighting each by its sample size. The value we obtain is .02258 ( $SD = .150$ ). We can also compute the variance expected solely from sampling error. The formula for the sampling error variance of each individual correlation  $r_j$  is

$$(1 - .331^2)^2 / (N_i - 1),$$

where .331 is the sample size-weighted mean of the correlations in [Table 1.1](#). If we weight each of these estimates by its sample size (as we did when we computed the observed variance), the formula for variance expected from sampling error is

$$S_e^2 = \frac{\sum_{i=1}^{i=30} \left[ \frac{N_i (1 - .331^2)^2}{N_i - 1} \right]}{\sum N_i}.$$

This value is .02058 ( $SD = .144$ ). The ratio of variance expected from sampling error to actual (observed) variance is .02058/.02258 = .91. Thus, sampling error alone accounts for an estimated 91% of the observed variance in the correlations. The square root of the .91 is the correlation between the sampling errors and the observed correlations. This correlation is .95 and is a more informative index than the percent of variance accounted for (Schmidt, 2010). The best conclusion is that the relationship between job satisfaction and or-

ganizational commitment is constant across sexes, races, job levels, ages, geographical locations, and size of organization. (The difference between a correlation of 1.00 and our value of .95 is due to second-order sampling error, which is discussed in [Chapter 9](#).) The best estimate of this constant value is .331, the sample size-weighted mean of the 30 correlations. When in our oral presentations researchers analyzed the data from these 30 studies qualitatively, different people came to different conclusions. In contrast, all researchers applying the quantitative method used here would (barring computational errors) come to exactly the same conclusion.

For theoretical purposes, the value .331 is not the one we want, because it is biased downward by unreliability in both measures. The effect of measurement error is to reduce all the observed correlations, and hence the mean correlation, below the actual correlation between the two constructs. What we are interested in is the construct-level correlation, because this correlation reflects the underlying science. Suppose from information in the 30 studies we estimate the average reliability of job satisfaction measures at .70 and the average reliability of organizational commitment measures at .60. Then the estimated correlation between true scores on the measures is  $.331 / \sqrt{.70(.60)} = .51$ . This value is the best estimate of the construct-level correlation. Schmidt, Le, and Oh (in press) have shown that true scores and construct scores typically correlate about .98, so true score correlations are good estimates of construct correlations. The necessity of correcting for measurement error is discussed by Hedges (2009b, chap. 3).

Most artifacts other than sampling error that distort study findings are systematic rather than random. They usually create a downward bias in the obtained study  $r$  or  $d$  value. For example, all variables in a study must be measured and all measures of variables contain measurement error. There are no exceptions to this rule. The effect of measurement error is to downwardly bias every correlation or  $d$  value. Measurement error can also cause *differences* between studies: If the measures used in one study have more measurement error than those used in another study, the observed  $r$ s or  $d$ s will be smaller in the first study. Thus, meta-analysis must correct both for the downward bias and for the artifactually created differences between different studies. Corrections of this sort are discussed in [Chapters 2 to 7](#).

Traditional review procedures are inadequate to integrate conflicting findings across large numbers of studies. As Glass (1976) pointed out, the results of hundreds of studies “can no more be grasped in our traditional narrative discursive review than one can grasp the sense of 500 test scores without the aid of techniques for organizing, depicting and interpreting data” (p. 4). In such areas as the effects of class size on student learning, the relationship of IQ to creativity, and the effects of psychotherapy on patients, literally hundreds of

studies can accumulate over a period of only a few years. Glass (1976) noted that such studies collectively contain much more information than can be extracted from them using narrative review methods. He pointed out that because we have not exploited these gold mines of information, “We know much less than we have proven.” What is needed are methods that will integrate results from existing studies to reveal patterns of relatively invariant underlying relationships and causalities, the establishment of which will constitute general principles and cumulative knowledge.

At one time in the history of psychology and the social sciences, the pressing need was for more empirical studies examining the problem in question. In many areas of research, the need today is not additional empirical data but some means of making sense of the vast amounts of data that have been accumulated. Because of the increasing number of areas within psychology and the other social sciences in which the number of available studies is quite large and the importance to theory development and practical problem solving of integrating conflicting findings to establish general knowledge, meta-analysis has come to play an increasingly important role in research. Such methods can be built around statistical and psychometric procedures that are already familiar to us. As Glass (1976) stated,

Most of us were trained to analyze complex relationships among variables in the primary analysis of research data. But at the higher level, where variance, nonuniformity and uncertainty are no less evident, we too often substitute literary exposition for quantitative rigor. The proper integration of research requires the same statistical methods that are applied in primary data analysis. (p. 6)

---

## Role of Meta-Analysis in the Behavioral and Social Sciences

The small-sample studies typical of psychological research produce seemingly contradictory results, and reliance on statistical significance tests causes study results to appear even more conflicting. Meta-analysis integrates the findings across such studies to reveal the simpler patterns of relationships that underlie the research literature, thus providing a basis for theory development. Meta-analysis can correct for the distorting effects of sampling error, measurement error, and other artifacts that produce the illusion of conflicting findings.

The goal in any science is the production of cumulative knowledge. Ultimately, this means the development

of theories that explain the phenomena that are the focus of the scientific area. One example would be theories that explain how personality traits develop in children and adults over time and how these traits affect their lives. Another would be theories of what factors cause job and career satisfaction and what effects job satisfaction in turn has on other aspects of one's life. Before theories can be developed, however, we need to be able to precisely calibrate the relationships between variables. For example, what is the relationship between peer socialization and level of extroversion? What is the relationship between job satisfaction and job performance?

Unless we can precisely calibrate such relationships among variables, we do not have the raw materials out of which to construct theories. There is nothing for a theory to explain. For example, if the relationship between extroversion and popularity of children varies capriciously across different studies from a strong positive to a strong negative correlation and everything in between, we cannot begin to construct a theory of how extroversion might affect popularity. The same applies to the relationship between job satisfaction and job performance.

The unfortunate fact is that most research literatures do show conflicting findings of this sort. Some studies find statistically significant relationships and some do not. In much of the research literature, this split is approximately 50–50 (Cohen, 1962, 1988; Schmidt, 2010; Schmidt & Hunter, 1997; Schmidt, Hunter, & Urry, 1976; Sedlmeier & Gigerenzer, 1989). This has been the traditional situation in most areas of the behavioral and social sciences. Hence, it has been very difficult to develop understanding, theories, and cumulative knowledge.

Today meta-analysis is being widely applied to solve this problem. The extent of the use of meta-analysis is mirrored in the fact that a Google search using this term produces over 50 million hits.

### **The Myth of the Perfect Study**

Before meta-analysis, the usual way in which scientists attempted to make sense of the research literature was by use of the narrative subjective review. In much of the research literature, however, there were not only conflicting findings but also large numbers of studies. This combination made the standard narrative subjective review a nearly impossible task—one shown by research on human information processing to be far beyond human capabilities. How does one sit down and make sense of, say, 210 conflicting studies?

The answer as developed in many narrative reviews was what came to be called the myth of the perfect

study. Reviewers convinced themselves that most—usually the vast majority—of the available studies were “methodologically deficient” and should not even be considered in the review. These judgments of methodological deficiency were often based on idiosyncratic ideas: One reviewer might regard the Peabody Personality Inventory as “lacking in construct validity” and throw out all studies that used that instrument. Another might regard use of that same inventory as a prerequisite for methodological soundness and eliminate all studies *not* using this inventory. Thus, any given reviewer could eliminate from consideration all but a few studies and perhaps narrow the number of studies from 210 to, say, 7. Conclusions would then be based on these seven studies.

It has long been the case that the most widely read literature reviews are those appearing in textbooks. The function of textbooks, especially advanced-level textbooks, is to summarize what is known in a given field. No textbook, however, can cite and discuss 210 studies on a single relationship. Textbook authors would often pick out what they considered to be the one or two “best” studies and then base textbook conclusions on just those studies, discarding the vast bulk of the information in the research literature. Hence, the myth of the perfect study.

In fact, there are no perfect studies. All studies contain measurement error in all measures used, as discussed later. Independent of measurement error, no study's measures have perfect construct validity. Furthermore, there are typically other artifacts that distort study findings. Even if a hypothetical (and it would have to be hypothetical) study suffered from none of these distortions, it would still contain sampling error—typically a substantial amount of sampling error—because sample sizes are rarely very large. Hence, no single study or small selected subgroup of studies can provide an optimal basis for scientific conclusions about cumulative knowledge. As a result, reliance on “best studies” did not provide a solution to the problem of conflicting research findings. This procedure did not even successfully deceive researchers into believing it was a solution—because different narrative reviewers arrived at different conclusions because they selected a different subset of “best” studies. Hence, the “conflicts in the literature” became “conflicts between the reviews.”

### **Some Relevant History**

By the mid-1970s, the behavioral and social sciences were in serious trouble. Large numbers of studies had accumulated on many questions that were important to theory development and/or social policy decisions.

Results of different studies on the same question typically were conflicting. For example, are workers more productive when they are satisfied with their jobs? The studies did not agree. Do students learn more when class sizes are smaller? Research findings were conflicting. Does participative decision making in management increase productivity? Does job enlargement increase job satisfaction and output? Does psychotherapy really help people? The studies were in conflict. As a consequence, the public and government officials were becoming increasingly disillusioned with the behavioral and social sciences, and it was becoming more and more difficult to obtain funding for research. In an invited address to the American Psychological Association in 1970, Senator Walter Mondale expressed his frustration with this situation:

What I have *not* learned is what we should do about these problems. I had hoped to find research to support or to conclusively oppose my belief that quality integrated education is the most promising approach. But I have found very little conclusive evidence. For every study, statistical or theoretical, that contains a proposed solution or recommendation, there is always another, equally well documented, challenging the assumptions or conclusions of the first. No one seems to agree with anyone else's approach. But more distressing I must confess, I stand with my colleagues confused and often disheartened.

Then, in 1981, the director of the Federal Office of Management and Budget, David Stockman, proposed an 80% reduction in federal funding for research in the behavioral and social sciences. This proposal was politically motivated in part, but the failure of behavioral and social science research to be cumulative created the vulnerability to political attack. This proposed cut was a trial balloon sent up to see how much political opposition it would arouse. Even when proposed cuts are much smaller than a draconian 80%, constituencies can usually be counted on to come forward and protest the proposed cuts. This usually happens, and many behavioral and social scientists expected it to happen. But it did not. The behavioral and social sciences, it turned out, had no constituency among the public; the public did not care (see "Cuts Raise New Social Science Query," 1981). Finally, out of desperation, the American Psychological Association took the lead in forming the Consortium of Social Science Associations to lobby against the proposed cuts. Although this super association had some success in getting these cuts reduced (and even, in some areas, getting increases in research funding in subsequent years), these developments should make us look carefully at how such a thing could happen.

The sequence of events that led to this state of affairs was much the same in one research area after another.



First, there was initial optimism about using social science research to answer socially important questions. Do government-sponsored job-training programs work? We will do studies to find out. Does Head Start really help disadvantaged kids? The studies will tell us. Does integration increase the school achievement of black children? Research will provide the answer. Next, several studies on the question are conducted, but the results are conflicting. There is some disappointment that the question has not been answered, but policy makers—and people in general—are still optimistic. They, along with the researchers, conclude that more research is needed to identify the supposed interactions (moderators) that have caused the conflicting findings. For example, perhaps whether job training works depends on the age and education of the trainees. Maybe smaller classes in the schools are beneficial only for lower IQ children. It is hypothesized that psychotherapy works for middle-class but not working-class patients. That is, the conclusion at this point is that a search for moderator variables is needed.

In the third phase, a large number of research studies are funded and conducted to test these moderator hypotheses. When they are completed, there is now a large body of studies, but instead of being resolved, the number of conflicts increases. The moderator hypotheses from the initial studies are not borne out, and no one can make sense out of the conflicting findings. Researchers conclude that the question that was selected for study in this particular case has turned out to be hopelessly complex. They then turn to the investigation of another question, hoping that this time the question will turn out to be more tractable. Research sponsors, government officials, and the public become disenchanted and cynical. Research funding agencies cut money for research in this area and in related areas. After this cycle has been repeated enough times, social and behavioral scientists themselves become cynical about the value of their own work, and they publish articles endorsing the belief that behavioral and social science research is incapable, *in principle*, of developing cumulative knowledge and providing general answers to socially important questions. Examples of this include Cronbach (1975), Gergen (1982), and Meehl (1978).

Clearly, at this point, there was a critical need for some means of making sense of the vast number of accumulated study findings. Starting in the late 1970s, new methods of combining findings across studies on the same subject were developed. These methods were referred to collectively as *meta-analysis*, a term coined by Glass (1976). Applications of meta-analysis to the accumulated research literature (e.g., Schmidt & Hunter, 1977) showed that research findings were not nearly as conflicting as had been thought and that useful and sound general conclusions could, in fact, be drawn from existing research. The conclusion was that cumulative theoretical knowledge is possible in the behavioral and social sciences, and socially important questions

can be answered in reasonably definitive ways. As a result, the gloom and cynicism that had enveloped many in the behavioral and social sciences lifted.

In fact, meta-analysis has even produced evidence that cumulativeness of research findings in the behavioral sciences is probably as great as that in the physical sciences. We have long assumed that our research studies are less consistent than those in the physical sciences. Hedges (1987) used meta-analysis methods to examine variability of findings across studies in 13 research areas in particle physics and 13 research areas in psychology. Contrary to common belief, his findings showed that there is as much variability across studies in physics as there is in psychology. Furthermore, he found that the physical sciences used methods to combine findings across studies that were “essentially identical” to meta-analysis. The research literature in both areas—psychology and physics—yielded cumulative knowledge when meta-analysis was properly applied. Hedges's major finding is that the frequency of conflicting research findings is no greater in the behavioral and social sciences than in the physical sciences. The fact that this finding has been so surprising to many social scientists points up the fact that we have long overestimated the consistency of research findings in the physical sciences. Furthermore, in the physical sciences, no research question can be answered by a single study, and physical scientists must and do use meta-analysis to make sense of their research literature, just as we do. (And, as noted earlier, in analyzing data in individual studies, the physical sciences do not use significance tests; they use point estimates and confidence intervals.)

Other changes have also been produced by meta-analysis. The relative status of reviews has changed dramatically. Journals that traditionally published only primary studies and refused to publish reviews have now published meta-analytic reviews in large numbers for some years. In the past, research reviews were based on the narrative subjective method, and they had limited status and gained little credit for one in academic raises or promotions. The rewards went to those who conducted primary research. Not only is this no longer the case, but there also has been a more important development. Today, many discoveries and advances in cumulative knowledge are being made not by those who do primary research studies but by those who use meta-analysis to discover the latent meaning of existing research literatures. Today, behavioral or social scientists with the needed training and skills are making major original discoveries and contributions by mining the untapped veins of information in the accumulated research literatures.

The meta-analytic process of cleaning up and making sense of the research literature not only reveals the cumulative knowledge that is there but also provides clearer directions about what the remaining research

needs are. That is, we also learn what kinds of primary research studies are needed next. However, some have raised the concern that meta-analysis may be killing the motivation and incentive to conduct primary research studies. Meta-analysis has clearly shown that no single primary study can ever resolve an issue or answer a question. Research findings are inherently probabilistic (Taveggia, 1974), and, therefore, the results of any single study could have occurred by chance. Only meta-analytic integration of findings across studies can control sampling error and other artifacts and provide a foundation for conclusions. And yet meta-analysis is not possible unless the needed primary studies are conducted. In new research areas, this potential problem is not of much concern. The first study conducted on a question contains 100% of the available research information. The second contains roughly 50%, and so on. Thus, the early studies in any area have a certain status. The 50th study, however, contains only about 2% of the available information, and the 100th, about 1%. Will we have difficulty motivating researchers to conduct the 50th or 100th study? If we do, we do not believe this will be due to meta-analysis. When the narrative review was the dominant method of research integration, reviewers did not base their conclusions on single studies but on multiple studies. So no researcher could reasonably hope then—as now—that his or her single study could decide an issue. In fact, meta-analysis represents an improvement for the primary researcher in one respect: All available relevant studies are included in a meta-analysis, and hence, every study has an effect. As we saw earlier, narrative reviewers often threw out most of the relevant studies and based their conclusions on a handful of their favorite studies.

Also, it should be noted that those who raise this question overlook a beneficial effect that meta-analysis has had: It prevents the diversion of valuable research resources into truly unneeded research studies. Meta-analysis applications have revealed that there are questions on which additional research would waste scientifically and socially valuable resources. For example, already as of 1980, 882 studies based on a total sample of 70,935 had been conducted relating measures of perceptual speed to the job performance of clerical workers. Based on these studies, our meta-analytic estimate of this mean correlation is .47 and its  $SD_p = .22$ ; Pearlman, Schmidt, & Hunter, 1980). For other abilities, there were often 200 to 300 cumulative studies. Clearly, further research on these relationships is not the best use of available resources.

If one or more new studies appear after a meta-analysis has been completed, how should the meta-analysis be updated to include these new studies? Schmidt and Raju (2007) showed that the optimal method is to recompute the meta-analysis including those studies (i.e., use the “medical method” of updating a meta-analysis rather than a Bayesian approach).

---

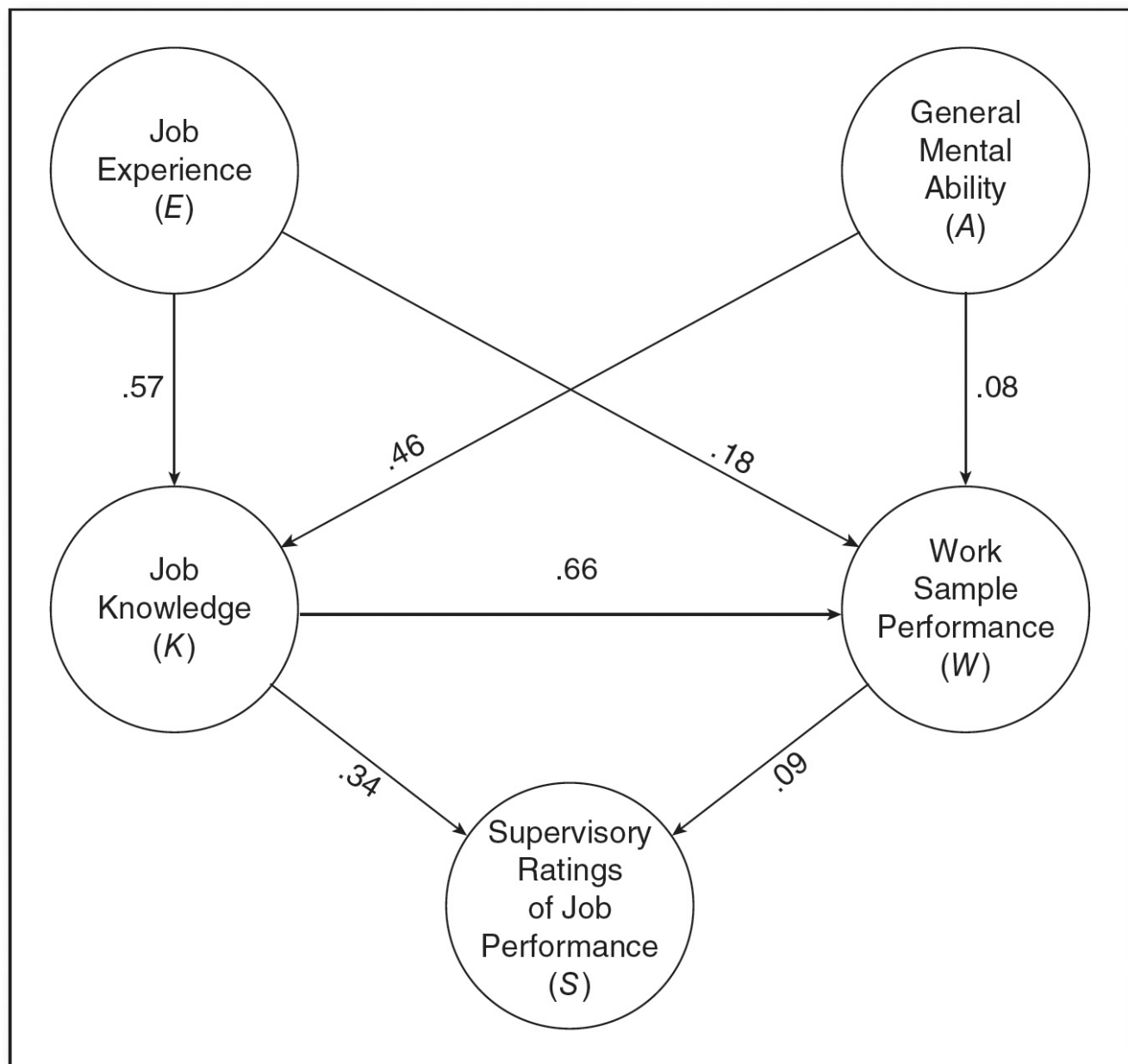
## Role of Meta-Analysis in Theory Development

As noted earlier, the major task in the behavioral and social sciences, as in other sciences, is the development of theory. A good theory is simply a good explanation of the processes that actually take place in a phenomenon. For example, what actually happens when employees develop a high level of organizational commitment? Does job satisfaction develop first and then cause the development of commitment? If so, what causes job satisfaction to develop and how does it affect commitment? How do higher levels of mental ability cause higher levels of job performance? Only by increasing job knowledge? Or also by directly improving problem solving on the job? The social scientist is essentially a detective; his or her job is to find out why and how things happen the way they do. To construct theories, however, we must first know some of the basic facts, such as the empirical relationships among variables. These relationships are the building blocks of theory. For example, if we know there is a high and consistent positive population correlation between job satisfaction and organization commitment, this will send us in particular directions in developing our theory. If the correlation between these variables is very low and consistent, theory development will branch in different directions. If the relationship is highly variable across organizations and settings, we will be encouraged to advance interactive or moderator-based theories. Meta-analysis provides these empirical building blocks for theory. Meta-analytic findings tell us what it is that needs to be explained by the theory. Meta-analysis has been criticized because it does not directly generate or develop theory (Guzzo, Jackson, & Katzell, 1986). This is akin to criticizing word processors because they do not generate books on their own. The results of meta-analysis are indispensable for theory construction, but theory construction itself is a creative process distinct from meta-analysis.

As implied in the language used in our discussion, theories are causal explanations. The goal in every science is explanation, and explanation is always causal. In the behavioral and social sciences, the methods of path analysis (see, e.g., Hunter & Gerbing, 1982) and structural equation modeling (SEM) can be used to test causal theories when the data meet the assumptions of the method. The relationships revealed by meta-analysis—the empirical building blocks for theory—can be used in path analysis and SEM to test causal theories. Experimentally determined relationships can also be entered into path analyses along with observationally based relationships. It is only necessary to transform  $d$  values to correlations (see [Chapter 7](#)). Thus, path analyses can be “mixed.” Path analysis and SEM cannot demonstrate that a theory is correct but can disconfirm a theory, that is, show that it is not correct. Path analysis can therefore be a powerful tool for reducing the number of theories that could possibly be consistent with the data, sometimes to a very small number, and

sometimes to only one theory (Hunter, 1988). For an example, see Hunter (1983a). Every such reduction in the number of possible theories is an advance in understanding.

Application of path analysis or SEM requires either the correlations among the theoretically relevant variables (correlation matrix) or the covariances among the variables (variance-covariance matrix). Meta-analysis can be used to create correlation matrices for the variables of interest. Because each meta-analysis can estimate a different cell in the correlation matrix, it is possible to assemble the complete correlation matrix, even though no single study has included every one of the variables of interest (see, e.g., Viswesvaran & Ones, 1995). If these correlations are appropriately corrected for the biasing effects of artifacts such as measurement error, it is then possible to apply path analysis or SEM to test causal (or explanatory) theories (Cook et al., 1992, pp. 315–316). One example of this is Schmidt, Hunter, and Outerbridge (1986). This study used meta-analysis to assemble the correlations among the variables of general mental ability, job experience, job knowledge, work sample performance, and supervisory evaluations of job performance. (These correlations were homogeneous across studies.) The path analysis results are shown in [Figure 1.1](#). This causal model fit the data quite well. As can be seen in [Figure 1.1](#), both job experience and general mental ability exert strong causal influence on the acquisition of job knowledge, which, in turn, is the major cause of high performance on the job sample measure. The results also indicate that supervisors based their ratings more heavily on employees' job knowledge than on actual performance capabilities. This causal model (or theory) of job performance has since been supported in other studies (Schmidt & Hunter, 1992). Today the research literature contains many studies that use meta-analysis in this manner.

**Figure 1.1 Path Model and Path Coefficients.**

*Note:* Adapted from “Impact of Job Experience and Ability on Job Knowledge, Work Sample Performance, and Supervisory Ratings of Job Performance,” by F. L. Schmidt, J. E. Hunter, and A. N. Outerbridge, 1986, *Journal of Applied Psychology*, 71, 432–439. Reprinted by permission of the authors.

Becker (1989; 2009, chap. 20) and Becker and Schram (1994) discussed the possibilities of using meta-analysis in this manner. Becker (1992) used this approach in examining a model of the variables affecting the achievement in science of male and female high school and college students. In that case, there were

insufficient studies available to obtain meta-analytic estimates of some of the needed correlations; nevertheless, progress was made and the information needed from future research was pinpointed. Becker (1996) and Shadish (1996) provided additional discussion. Although there are technical complexities that must be dealt with in using meta-analysis in this manner (Cook et al., 1992, pp. 328–330), it is a promising approach to accelerating cumulative knowledge in the social sciences. Technical questions related to this use of meta-analysis are discussed in [Chapter 5](#).

---

## Meta-Analysis in Industrial-Organizational Psychology

There have been numerous applications of meta-analysis in industrial-organizational (I/O) psychology. The most extensive and detailed application of meta-analysis in I/O psychology has been the study of the generalizability of the validities of employment selection procedures (Schmidt, 1988; Schmidt & Hunter, 1981, 1998). The findings have resulted in major changes in the field of personnel selection. Validity generalization research is described in more detail in [Chapter 4](#).

Recent meta-analyses in organizational psychology have addressed a broad range of topics across different levels of analysis. At the levels of the organization or business unit, Harter, Schmidt, and Hayes (2002); Harter, Schmidt, Asplund, and Kilham (2010); and Whitman, Van Rooy, and Viswesvaran (2010) demonstrated that unit-level job satisfaction and employee engagement has positive, generalized effects on business unit financial performance and customer satisfaction. Another meta-analysis showed, across 83 different organizations, the generalized positive effects on job performance of the Productivity Measurement and Enhancement System (ProMES; Pritchard, Harrell, DiazGranadaos, & Guzman, 2008), a performance management system designed by organizational psychologists to provide workers and employees with quick and effective feedback on their performance. Meta-analyses of team research continue to be popular, with one meta-analysis summarizing how different teamwork processes affect team effectiveness (LePine, Piccolo, Jackson, Mathieu, & Saul, 2008).

Other meta-analyses focused on individuals as the unit of analysis. One such study examined the relationship between job turnover and the five-factor model (FFM) of personality and found that Emotional Stability is a generalized negative predictor of turnover (Zimmerman, 2008). Another study attempted to untangle the ambiguous causal relationship between job attitudes and job performance (Riketta, 2008). Other studies focused on multicultural and international issues. Dean, Roth, and Bobko (2008) meta-analytically examined ethnic



and gender subgroup differences in assessment center ratings to show that gender differences in ratings from these evaluations are smaller than previously thought, but that some ethnic subgroup differences are larger than previously believed. Taras, Kirkman, and Steel (2010) showed that Hofstede's (1980) cultural value dimensions had validity in predicting (in decreasing order) individual emotions, attitudes, behaviors, and job performance. Geyskens, Krishnan, Steenkamp, and Cunha (2009) presented an extensive examination of the use of meta-analysis in management-related research.

Older examples also span a variety of topics and units of analysis. C. D. Fisher and Gitelson's (1983) meta-analysis examined the negative and positive correlates of conflict and ambiguity for members' roles in teams. Meta-analyses of leadership performance were also popular. An example is a meta-analytic test of Fiedler's contingency theory of leadership, a dominant theory of leadership at the time (L. H. Peters, Harthe, & Pohlman, 1985). Other meta-analyses studied questions about attitudes and beliefs, such as the relatively low accuracy of self-ratings of ability and skill (Mabe & West, 1982) and the negative relationship between job satisfaction and absenteeism (Terborg & Lee, 1982). Other studies focused on more specific interventions and assessments, such as the small but positive effect of realistic job previews in reducing subsequent employee turnover (Premack & Wanous, 1985); the positive, generalizable validity of the LSAT for predicting performance in law school (Linn, Harnisch, & Dunbar, 1981a); and the limited abilities of financial analysts to predict stock growth (Coggin & Hunter, 1983). In short, researchers have pursued and continue to pursue meta-analytic studies across a wide variety of subjects and continue to recognize psychometric meta-analysis as an important research tool.

Additional examples of influential meta-analyses are found in the literature on managerial training. Burke and Day's (1986) meta-analysis on the effectiveness of managerial training prompted a subsequent stream of meta-analytic research on management training, including D. B. Collins and Holton (2004); Taylor, Russ-Eft, and Taylor (2009); and Powell and Yalcin (2010). The conclusions from these studies have repeatedly been that management training programs can be effective in nearly all situations at changing particular behaviors and for the acquisition of knowledge, particularly in the areas of time management and human relations skills. Other meta-analyses have assessed the results of training across a variety of organizational contexts. For example, a recent meta-analysis examined how combinations of training content, trainee attributes, and trainees' affective reactions to training influence the outcomes of a training program (Sitzmann, Brown, Casper, Ely, & Zimmerman, 2008). Other meta-analytic studies investigated the relationships among different training criteria such as behavior, learning, and performance (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997), or how trainees use their training in applied settings and share knowledge from training



with others (Arthur, Bennett, Edens, & Bell, 2003). All three meta-analyses have had powerful impacts on traditional models of learning and training in I/O psychology, resulting in updates to D. L. Kirkpatrick's (2000) widely used model of learning, a demonstration of the effectiveness of lectures as a form of training, and reconsideration of the value of affective reactions to training.

Colquitt, LePine, and Noe (2000) used meta-analysis to further expand on Alliger et al.'s (1997) work and examine how individual differences, situational factors, and job career factors influenced an individual's motivation during training and the subsequent outcomes of this motivation. Colquitt et al.'s (2000) work established the importance of motivation as a critical factor in determining the efficacy of training, how individuals transfer learning from training to their work performance, and how effective they are in sharing training knowledge with others. In a more recent study of the impact of motivation, Payne, Youngcourt, and Beaubien (2007) have used meta-analytic methods to calibrate the impact of goal orientation (a psychological variable reflecting the motivation of individuals to learn versus the motivation to perform well in front of others) on training outcomes. These two meta-analyses have influenced how researchers conduct subsequent primary studies and created new opportunities for researchers to examine the role of motivation in training. In short, meta-analysis has affected research on management training, motivation in training, and training in general by demonstrating the validity and value of training to organizations across the board.

Meta-analysis has also been used extensively in the leadership literature, a popular topic of study in I/O psychology and organizational behavior (OB). Meta-analysis has provided some clarity on a difficult subject, as well as changing how studies were conducted in light of the findings of the meta-analyses. Judge, Colbert, and Ilies (2004) reported a relatively low relationship between general cognitive ability and leadership, noting that this relationship is weaker than expected based on earlier qualitative reviews. Similar work was done demonstrating that leadership cannot be explained solely as a result of personality traits in the popular FFM framework (Judge, Bono, Ilies, & Gerhardt, 2002). Several years later, this line of research inspired a group of researchers to consider conceptual models of positive ("bright-side") and negative ("dark-side") personality traits in leaders linked to leadership emergence and efficacy (Judge, Piccolo & Kosalka, 2009). Finally, multiple meta-analyses have been used to show the validity, uniqueness, and importance of transformational (charismatic) leadership, a specific set of leadership behaviors highly motivating to employees and a central topic of research in the leadership literature (Bono & Judge, 2004; Eagly, Johannsen-Schmidt, & van Engen, 2003; Eagly, Karau, & Makhijani, 1995). Applications of meta-analysis have produced important changes in research conclusions, fundamental changes in existing theoretical paradigms, and the development of new lines of research.

---

## Wider Impact of Meta-Analysis on Psychology

Some have viewed meta-analysis as merely a set of improved methods for doing literature reviews. Meta-analysis is actually more than that. By quantitatively comparing findings across diverse studies, meta-analysis can discover new knowledge not inferable from any individual study and can sometimes answer questions that were never addressed in any of the individual studies contained in the meta-analysis. For example, no individual study may have compared the effectiveness of a training program for people of higher and lower mental ability; by comparing mean  $d$  value statistics across different groups of studies, however, meta-analysis can reveal this difference. That is, moderator variables (interactions) never studied in any individual study can be revealed by meta-analysis, greatly facilitating the development of cumulative knowledge. M. E. Chan and Arvey (2012) provide an analysis of the positive impact of meta-analysis on the development of knowledge in psychology and the social sciences. Richard, Bond, and Stokes-Zoota (2003) review the impact of 322 meta-analyses in social psychology. And Dieckmann, Malle, and Bodner (2009) provide an assessment of the use of meta-analysis in several areas of psychological research. Carlson and Ji (2011) presented an analysis of how meta-analytic studies are used and cited in the wider psychological literature. They found that the frequency of citations to meta-analyses (vs. primary studies) has been rapidly increasing in recent years.

Even though it is much more than that, meta-analysis is indeed an improved method for synthesizing or integrating the research literature. The premier review journal in psychology is *Psychological Bulletin*. Over time since 1980, a steadily increasing percentage of the reviews published in this journal have been meta-analyses and a steadily decreasing percentage have been traditional narrative subjective reviews. It is not uncommon for narrative review manuscripts to be returned by editors to the authors with the request that meta-analysis be applied to the studies reviewed (Cooper, 2003). Most of the remaining narrative reviews published today in *Psychological Bulletin* focus on research literature that is not well enough developed to be amenable to quantitative treatment.

Although a movement toward change began about 2010, most of the meta-analyses that have appeared in *Psychological Bulletin* have employed fixed effects methods, resulting in many cases in overstatement of the precision of the meta-analysis findings (Schmidt, Oh, & Hayes, 2009). (See [Chapters 5](#) and [8](#) for a discussion of fixed vs. random meta-analysis models; confidence intervals are too narrow when the fixed effect model is used.) Despite this fact, these meta-analyses produced findings and conclusions that are more accurate than those produced by the traditional narrative subjective method. Many other journals have shown the same in-

crease over time in the number of meta-analyses published. Many of these journals had traditionally published only individual empirical studies and had rarely published reviews up until the advent of meta-analysis in the late 1970s. These journals began publishing meta-analyses because meta-analyses came to be viewed not as “mere reviews” but as a form of empirical research. As a result of this change, the quality and accuracy of conclusions from the research literature improved in a wide variety of journals and in a corresponding variety of research areas in psychology. This improvement in the quality of conclusions from the research literature has expedited theory development in many areas in psychology.

The impact of meta-analysis on psychology textbooks has been positive and dramatic. Textbooks are important because their function is to summarize the state of cumulative knowledge in a given field. Most people—students and others—acquire most of their knowledge about psychological theory and findings from their reading of textbooks. Prior to meta-analysis, textbook authors faced with hundreds of conflicting studies on a single question subjectively and arbitrarily selected a small number of their preferred studies from the literature and based the textbook conclusions on only those few studies. Today, most textbook authors base their conclusions on meta-analysis findings (Myers, 1991), making their conclusions and their textbooks much more accurate. We cannot overemphasize the importance of this development in advancing cumulative knowledge in psychology.

Because multiple studies are needed to solve the problem of sampling error, it is critical to ensure the availability of all studies on each topic. A major problem is that many good replication articles are rejected by the primary research journals. Journals currently put excessive weight on surprising and novel findings in evaluating studies and often fail to consider either sampling error or other important technical problems such as measurement error. Many journals will not even consider “mere replication studies” or “mere measurement studies.” Many persistent authors eventually publish such studies in journals with lower prestige, but they must endure many letters of rejection, and publication is often delayed for a long period of time. Problems of this sort, including the general issue of publication bias, are discussed in more detail in [Chapter 13](#).

To us, this clearly indicates that we need a new type of journal—whether hard copy or electronic—that systematically archives all studies that will be needed for later meta-analyses. The American Psychological Association's Experimental Publication System in the early 1970s was an attempt in this direction. However, at that time, the need subsequently created by meta-analysis did not yet exist; the system apparently met no real need at that time and hence was discontinued. Today, the need is so great that failure to have such a journal system in place is retarding our efforts to reach our full potential in creating cumulative knowledge in

psychology and the social sciences.

In view of the large number of meta-analyses available in the psychology and social sciences literatures, some readers may wonder why the examples we use in this book to illustrate meta-analysis principles and methods do not employ data from those meta-analyses. The primary reason is that the amount of data (the number of correlations or  $d$  statistics) is usually so large as to result in cumbersome examples. For pedagogical reasons, we have generally employed examples consisting of small numbers of studies in which the data are hypothetical. As explained in the following chapters, meta-analyses based on such small numbers of studies would not ordinarily yield results that would be optimally stable. (We discuss second-order sampling error in [Chapter 9](#).) However, such examples provide the means to simply and clearly illustrate the principles and methods of meta-analysis, and we believe this is the crucial consideration.

---

## Impact of Meta-Analysis outside Psychology

### Impact in Biomedical Research

The impact of meta-analysis may be even greater in biomedical research than in the behavioral and social sciences (Hunt, 1997, chap. 4). Hundreds of meta-analyses have been published in leading medical research journals such as the *New England Journal of Medicine* and the *Journal of the American Medical Association*. Already as of 1995, the medical literature contained between 962 and 1,411 meta-analyses, depending on the method of counting (Moher & Olkin, 1995). This number is much greater today. In medical research, the preferred study is the randomized controlled trial (RCT), in which participants are assigned randomly to receive either the treatment or a placebo, with the researchers being blind as to which treatment the participants are receiving. Despite the strengths of this research design, it is usually the case that different RCTs on the same treatment obtain conflicting results. This is partly because the effect sizes are often small and partly because (contrary perhaps to widespread perceptions) RCTs are often based on small sample sizes. In addition, the problem of information overload is even greater in medicine than in the social sciences; more than a million medical research studies are published every year. No practitioner can possibly keep up with the medical literature in his or her area.

The leader in introducing meta-analysis to medical research was Thomas Chalmers. In addition to being a re-

searcher, Chalmers was also a practicing internal medicine physician who became frustrated with the inability of the vast, scattered, and unfocused medical research literature to provide guidance to practitioners. Starting in the mid-1970s, Chalmers developed his initial meta-analysis methods independently of those developed in the social and behavioral sciences. Despite being well conducted, his initial meta-analyses were not well accepted by medical researchers, who were critical of the concept of meta-analysis. In response, he and his associates developed “sequential meta-analysis”—a technique that reveals the date by which enough information had become available to show conclusively that a treatment was effective. Suppose, for example, that the first RCT for a particular drug had been conducted in 1975 but had a wide confidence interval, one that spans zero effect. Now, suppose three more studies had been conducted in 1976, providing a total of four studies to be meta-analyzed—and the confidence interval for the meta-analytic mean of these studies is still wide and still includes 0. Now, suppose five more RCTs had been conducted in 1977, providing nine studies for a meta-analysis up to this date. Now, if that meta-analysis yields a confidence interval that excludes 0, then we conclude that, given the use of meta-analysis, enough information was already available in 1977 to begin using this drug. On the basis of their meta-analysis findings and statistics on the disease, Chalmers and his associates then computed how many lives would have been saved to date had use of the drug begun in 1977. It turned out that, considered across different treatments, diseases, and areas of medical practice, a very large number of lives would have been saved had medical research historically relied on meta-analysis. The resulting article (Antman, Lau, Kupelnick, Mosteller, & Chalmers, 1992) is widely considered the most important and influential meta-analysis study ever published in medicine. It was even reported and discussed widely in the popular press (for example, the *New York Times* Science Section). It assured a major role for meta-analysis in medical research from that point on (Hunt, 1997, chap. 4).

Chalmers was also one of the driving forces behind the establishment of the Cochrane Collaboration, an organization that applies sequential meta-analysis in medical research in real time. This group conducts meta-analyses in a wide variety of medical research areas—and then updates each meta-analysis as new RCTs become available. That is, when a new RCT becomes available, the meta-analysis is rerun with the new RCT included. Hence, each meta-analysis is always current. The results of these updated meta-analyses are available on the Internet to researchers and medical practitioners around the world. It is likely that this effort has saved hundreds of thousands of lives by improving medical decision making. The Cochrane Collaboration website is [www.cochrane.org](http://www.cochrane.org). Another major early contributor to the development of meta-analysis methods for medical research is Richard Peto (1987). Later in the book, we discuss in detail methods of correcting for

the biasing effects of measurement error, using methods from psychometric theory. Completely independently of these methods from psychometric theory, Peto developed different but equivalent methods of correcting for measurement error in medical research. For an application of these methods, see MacMahon et al. (1990). Examples of the use in medical research of the methods presented in this book include Fountoulakis, Conda, Vieta, and Schmidt (2009) and Gardner, Frantz, and Schmidt (1999). Problems related to publication bias and research fraud in the biomedical sciences are discussed in [Chapter 13](#).

### Impact in other Disciplines

Meta-analysis has also become important in research in finance, marketing, sociology, ecology, and even wildlife management. Other areas in which meta-analysis is now being used include (Hafdahl, 2012) higher education, biological psychiatry, physical therapy, nursing practice, neuroscience, cardiology, forestry, and occupational counseling. An example of meta-analysis use in criminal justice is provided in Gendreau and Smith (2007). In fact, today it is difficult to find a research area in which meta-analysis is unknown. In the broad areas of education and social policy, the Campbell Collaboration is attempting to do for the social sciences what the Cochrane Collaboration (on which it is modeled) has done for medical practice (Rothstein, 2003; Rothstein, McDaniel, & Borenstein, 2001). Among the social sciences, perhaps the last to assign an important role to meta-analysis has been economics. However, meta-analysis has recently become important in economics, too (see, e.g., T. D. Stanley, 1998, 2001; T. D. Stanley & Jarrell, 1989, 1998). Another example in economics is the meta-analysis by Harmon, Oosterbeek, and Walker (2000); their study meta-analyzed a large number of studies on the financial returns to education and found an overall average rate of return of 6.5%. They also found that returns to education have fallen since the 1960s. There now is a doctoral program in meta-analysis in economics ([www.feweb.vu.nl/re/Master-Point/](http://www.feweb.vu.nl/re/Master-Point/)). In 2008, an international conference on meta-analysis usage in economics was held in Nancy, France (Nancy-Universite, 2008). Meta-analysis is now also used in political science (see, e.g., Pinello, 1999).

---

## Meta-Analysis and Social Policy

By providing the best available empirically based answers to socially important questions, meta-analysis can

influence public policy making (Hoffert, 1997; Hunter & Schmidt, 1996). This can be true for any public policy question for which there is a relevant research literature—which today includes most policy questions. Examples range from the Head Start program to binary chemical weapons (Hunt, 1997, chap. 6). The purpose of the Campbell Collaboration, described previously, is specifically to provide policy-relevant information to policy makers in governments and other organizations by applying meta-analysis to policy-relevant research literatures on social experiments. For more than 20 years, the U.S. General Accounting Office (GAO; now renamed the General Accountability Office), a research and evaluation arm of the U.S. Congress, has used meta-analysis to provide answers to questions posed by senators and representatives. For example, Hunt (1997, chap. 6) described how a GAO meta-analysis of the effects of the Women, Infants, and Children (WIC) program, a federal nutritional program for poor pregnant women, apparently changed the mind of Senator Jesse Helms and made him a supporter of the program. The meta-analysis found evidence that the program reduced the frequency of low-birth-weight babies by about 10%.

This meta-analysis was presented to Senator Helms by Eleanor Chelimsky, for years the director of the GAO's Division of Program Evaluation and Methodology. In that position, she pioneered the use of meta-analysis at GAO. Chelimsky (1994) stated that meta-analysis has proven to be an excellent way to provide Congress with the widest variety of research results that can hold up under close scrutiny under the time pressures imposed by Congress. She stated that the GAO has found that meta-analysis reveals both what is known and what is not known in a given topic area and distinguishes between fact and opinion “without being confrontational.” One application she cited as an example was a meta-analysis of studies on the merits of producing binary chemical weapons (nerve gas in which the two key ingredients are kept separate for safety until the gas is to be used). The meta-analysis did not support the production of such weapons. This was not what the Department of Defense (DOD) wanted to hear, and the DOD disputed the methodology and the results. The methodology held up under close scrutiny, however, and in the end Congress eliminated funds for these binary weapons.

By law, it is the responsibility of the GAO to provide policy-relevant research information to Congress. So the adoption of meta-analysis by the GAO is a clear example of the impact that meta-analysis can have on public policy. Although most policy decisions probably depend as much on political as on scientific considerations, it is possible for scientific considerations to have an impact with the aid of meta-analysis (Hoffert, 1997). Cordray and Morphy (2009) provide an extended discussion of the role of meta-analysis in the formulation of public policy. They emphasize the point that special care must be taken to ensure the objectivity of policy-related meta-analyses. In particular, they examine the meta-analysis conducted by the Environmental



Protection Agency on the health effects of secondhand smoke as an example of a biased meta-analysis. This example is important because numerous laws were passed against environmental tobacco smoke based on a faulty meta-analytic conclusion that it was harmful to health. Cordray and Morphy (2009) outline the steps necessary to ensure objectivity in meta-analyses.

---

## Meta-Analysis and Theories of Data and Theories of Knowledge

Every method of meta-analysis is of necessity based on a theory of data. It is this theory (or understanding of data) that determines the nature of the resulting meta-analysis methods. A complete theory of data includes an understanding of sampling error, measurement error, biased sampling (range restriction and range enhancement), dichotomization and its effects, data errors, and other causal factors that distort the raw data results we see in research studies. Once a theoretical understanding of how these factors affect data is developed, it becomes possible to develop methods for correcting for their effects. The necessity of doing so is presented in detail in Schmidt, Le, and Oh (2009). In the language of psychometrics, the first process—the process by which these factors (artifacts) influence data—is modeled as the attenuation model. The second process—the process of correcting for these artifact-induced biases—is called the disattenuation model. If the theory of data on which a method of meta-analysis is based is incomplete, that method will fail to correct for some or all of these artifacts and will thus produce biased results. For example, a theory of data that fails to recognize measurement error will lead to methods of meta-analysis that do not correct for measurement error. Such methods will then perforce produce biased meta-analysis results. Most current methods of meta-analysis do not, in fact, correct for measurement error, as noted in [Chapter 11](#). But in research methodology, the thrust is always in the direction of increased accuracy, so eventually methods for meta-analysis that do not correct for study artifacts that distort empirical findings will have to incorporate these corrections. This has already happened to some extent in that some users of these methods have “appended” these corrections to those methods (e.g., Aguinis, Sturman, & Pierce, 2008; Hall & Brannick, 2002). One's theory of data is also part of one's theory of knowledge or theory of epistemology. Epistemology is concerned with the ways in which we can attain correct knowledge. One requirement of an effective epistemology in empirical research is proper correction for the artifacts that distort the empirical data.

Sampling error and measurement error have a unique status among the statistical and measurement artifacts



with which meta-analysis must deal: They are *always* present in all real data. Other artifacts, such as range restriction, artificial dichotomization of continuous variables, or data transcription errors, may be absent in a particular set of studies being subjected to meta-analysis. There is always sampling error, however, because sample sizes are never infinite. Likewise, there is always measurement error, because there are no perfectly reliable measures. In fact, as we will see in subsequent chapters, it is the requirement of dealing simultaneously with both sampling error and measurement error that makes even relatively simple meta-analyses sometimes seem complicated. We are used to dealing with these two types of errors separately. For example, when psychometric texts (e.g., Lord & Novick, 1968; Nunnally & Bernstein, 1994) discuss measurement error, they assume an infinite (or very large) sample size, so that the focus of attention can be on measurement error alone, with no need to deal simultaneously with sampling error. Similarly, when statistics texts discuss sampling error, they implicitly assume perfect reliability (the absence of measurement error), so that they and the reader can focus solely on sampling error. Both assumptions are highly unrealistic because all real data simultaneously contain both types of errors. It is admittedly more complicated to deal with both types of errors simultaneously, yet this is what meta-analysis must do to be successful (see, e.g., Cook et al., 1992, pp. 315–316, 325–328; Matt & Cook, 2009, chap. 28).

The question of what theory of data (and therefore of knowledge) underlies a method of meta-analysis is strongly related to the question of what the general purpose of meta-analysis is. Glass (1976, 1977) stated that the purpose is simply to summarize and describe the studies in the research literature. As we will see in this book, our view (the alternative view) is that the purpose is to estimate as accurately as possible the construct-level relationships in the population (i.e., to estimate population values or parameters), because these are the relationships of scientific interest. This is an entirely different task; this is the task of estimating what the findings would have been if all studies had been conducted perfectly (i.e., with no methodological limitations). Doing this requires correction for sampling error, measurement error, and other artifacts (if present) that distort study results. Simply describing the contents of studies in the literature requires no such corrections and does not allow estimation of the parameters of scientific interest.

Rubin (1990, 1992) critiqued the common, descriptive concept of the purpose of meta-analysis and proposed the alternative offered in this book and previous editions of this book. He stated that, as scientists, we are not really interested in the population of imperfect studies per se, and hence an accurate description or summary of the results of these studies is not really important. Instead, he argued that the goal of meta-analysis should be to estimate the true effects or relationships—defined as “results that would be obtained in an infinitely large, perfectly designed study or sequence of such studies.” According to Rubin (1990),

Under this view, we really do not care *scientifically* about summarizing this finite population (of observed studies). We really care about the underlying scientific process—the underlying process that is generating these outcomes that we happen to see—that we, as fallible researchers, are trying to glimpse through the opaque window of imperfect empirical studies. (p. 157)

This is an excellent summary of the purpose of meta-analysis as we see it and as embodied in the methods presented in this book.

---

## Conclusion

Until recently, the psychological research literature was conflicting and contradictory. As the number of studies on each particular question became larger and larger, this situation became increasingly frustrating and intolerable. This situation stemmed from reliance on defective procedures for achieving cumulative knowledge: the statistical significance test in individual primary studies in combination with the narrative subjective review of the research literature. Meta-analysis principles have now correctly diagnosed this problem and have provided the solution. In area after area, meta-analytic findings have shown that there is much less conflict between different studies than had been believed; that coherent, useful, and generalizable conclusions can be drawn from the research literature; and that cumulative knowledge is possible in psychology and the social sciences. These methods have also been adopted in other areas such as medical research. A prominent medical researcher, Thomas Chalmers (as cited in Mann, 1990), has stated, “[Meta-analysis] is going to revolutionize how the sciences, especially medicine, handle data. And it is going to be the way many arguments will be ended” (p. 478). In concluding his oft-cited review of meta-analysis methods, Bangert-Drowns (1986) stated,

Meta-analysis is not a fad. It is rooted in the fundamental values of the scientific enterprise: replicability, quantification, causal and correlational analysis. Valuable information is needlessly scattered in individual studies. The ability of social scientists to deliver generalizable answers to basic questions of policy is too serious a concern to allow us to treat research integration lightly. The potential benefits of meta-analysis method seem enormous. (p. 398)

<https://doi.org/10.4135/9781483398105>

## Appendix Windows-Based Meta-Analysis Software Package Version 2.0

The Hunter-Schmidt Meta-Analysis Programs Package includes six programs that implement all basic types of Hunter-Schmidt psychometric meta-analysis methods. Information on how to obtain this program package is provided at the end of the Appendix. Brief descriptions of the six programs are provided in Section 4, “Types of Analyses.” A full description of the output of each program is given in Section 11, “Full Description of Output of Individual Programs.” These programs are intended to be used in conjunction with this book (henceforth referred to as “the text”).

In response to feedback from users, we have made a number of improvements to the current version of these programs (Version 2.0) that were not available in previous versions (Versions 1.0 and 1.1). These include the following:

1. The programs now allow importation of data files from Excel.
2. The programs now will produce forest plots. Forest plots can be useful in examining the studies prior to the meta-analysis.
3. Confidence intervals (in addition to credibility intervals) are now provided for all mean values except bare-bones meta-analysis means.
4. Output for corrected  $r$  and  $d$  values is now presented in tabular form, as well as in the more detailed output form. The tabular form output includes the values that should be reported in tables of meta-analytic results in papers submitted for publication. (Tabular form is not provided for bare-bones meta-analysis results.)
5. Studies can now be coded for analysis of potential moderators, making it more convenient to subgroup studies for moderator analysis.
6. The programs now include a method of checking for publication bias—cumulative meta-analysis. See [Chapter 13](#) for a discussion of this procedure.
7. Output now includes the correlation between observed  $r$  or  $d$  values and artifact effects on these values.
8. The maximum number of studies in a meta-analysis has been raised from 200 to 1,000.
9. It is now possible to select saved data files for analysis by just clicking on the file name. (In older versions, the file name had to be typed in.)
10. It is now easier for users to access the supplementary programs for computing composite

correlations and for converting point-biserial correlations to biserial correlations. These programs are no longer password protected. They can now be accessed by clicking on the appropriate icon.

11. The programs can be downloaded from the Internet. Alternatively, a program package CD can be provided.
12. Three small technical adjustments have been made. First, in the previous version, if an  $r$  or  $d$  value was entered as zero, the program would not run. This has been corrected. Second, the correction for the small bias in observed  $r$ s has been added. Third, in computing the sampling error variance of  $d$  values, the programs now use [Equation \(7.23a\)](#) instead of [Equation \(7.23\)](#). (These equations are in [Chapter 7](#).) Use of [Equation \(7.23a\)](#) produces greater accuracy in those cases in which the sample size for one group is very small compared to the other group. It is also accurate when the group sample sizes are less unequal or equal in size.

The programs are provided on a CD or downloaded from the Internet and are compatible with Microsoft Windows operating systems (Windows 95, 98, 98SE, ME, 2000, XP, and Windows 7). Program interface is logically and intuitively arranged so that people with basic familiarity with Windows-based applications can easily learn to use the program functions. Navigating through the different steps (pages) of the programs is achieved by single-clicking appropriate buttons or icons. Throughout all the steps (pages), there are built-in help functions in the form of roll-over pop-ups (i.e., help statements that appear when the cursor is rolled over certain predetermined areas) explaining the options and instructing the user in how to execute his or her desired tasks.

---

## What is in the Download or the CD Program Package?

In addition to the core meta-analysis programs, the program package contains (1) the Setup file (Setup Meta Program.exe), (2) the Readme.doc file (the file you are now reading), and (3) three utility programs in the form of Microsoft Excel templates (Composites.xls and Point-Biserial.xls; details of these programs are described in the “Extras” section). Two of these auxiliary programs help users convert correlations in the primary studies to appropriate forms before inputting them into the meta-analysis programs. The third auxiliary program produces forest plots of the studies in a meta-analysis.

---

## Installation

If you have an earlier version of the program package (Version 1.0 or 1.1), you should uninstall it before installing the current version. Uninstallation can easily be done by using the “Remove Hunter-Schmidt Meta Analysis Programs” option available in the Windows Programs taskbar.

The installation process starts by double-clicking the Setup file on the CD or following the downloading instructions you receive via an e-mail. You will be asked to provide a serial number to continue. The main programs and all the supporting files will be copied to folder “C: \ Meta Analysis Programs” on your hard drive (unless you specify a different drive).

---

## Starting the Program

The programs can be activated by selecting the “Hunter & Schmidt Meta Analysis Programs” in the Windows Programs taskbar. Alternatively, you can start the programs by using the icon “Hunter & Schmidt MA” on your desktop. You will be presented with the Start Page, where you can access the Readme file by clicking on the book icon at the upper right corner of the page. There are also links that allow you to get access to the extra programs for computing composite correlations or point-biserial correlations or for creating the forest plots. To move on to the next page, in which you can select types of meta-analyses to be run, click on the red arrow icon at the lower right corner of the page. The “Type of Analyses” page will appear, presenting you with four options of analyses (described next).

---

## Types of Analyses

The programs in this package do the following types of meta-analysis:

1. Meta-analysis of correlations corrected individually for the effects of artifacts. (These two programs are collectively referred to in the text as VG6.) These programs are used when (1) the user desires to estimate the correlation between variables, and (2) information on the statistical and measurement artifacts (i.e., range restriction, reliabilities on both variables) is

available in all (or the majority) of the primary studies. These programs are illustrated in [Chapter 3](#) of the text.

There are two subprograms under this type of meta-analysis:

- A. Program that corrects for direct range restriction: To be used when range restriction is direct (i.e., selection occurs on one of the two variables being correlated). (This program is referred to in the text as VG6-D; see [Chapter 3](#).)
- B. Program that corrects for indirect range restriction: To be used when range restriction is indirect (i.e., selection occurs on a third variable that is correlated with both the variables of interest). (This program is referred to in the text as VG6-I; see [Chapter 3](#).)

(Note: Both programs ask the user if there is any range restriction. If the answer is no, the data entry field for the range restriction statistic [the  $u$  value] does not appear. When there is no range restriction, the results provided by the subprograms 1A and 1B are identical.)

2. Meta-analysis of correlations using artifact distributions. These programs are used when (1) the user desires to estimate the correlation between variables, and (2) information on the statistical and measurement artifacts is *not* available in most of the primary studies. (These two programs are collectively referred to in the text as INTNL; see [Chapter 4](#).)

This meta-analysis program also includes two subprograms:

- A. Program that corrects for direct range restriction: To be used when range restriction is direct (i.e., selection occurs on one of the two variables being correlated). (This program is referred to in the text as INTNL-D; see [Chapter 4](#).)
- B. Program that corrects for indirect range restriction: To be used when range restriction is indirect (i.e., selection occurs on a third variable that is correlated with both the variables of interest). (This program is referred to in the text as INTNL-I; see [Chapter 4](#).)

(Note: Both programs ask the user if there is any range restriction. If the answer is no, the data entry field for the range restriction statistic [the  $u$  value] does not appear. When there is no range restriction, the results provided by the subprograms 2A and 2B are identical.)

3. Meta-analysis of  $d$  values corrected individually for measurement error.

This program is used when (1) the meta-analysis is based on the effect size ( $d$  value; standardized difference between groups), and (2) information on the reliability of the measure of the dependent variable is available in all (or most) of the primary studies. (This program is referred to in the text as D-VALUE; see [Chapter 7](#).)

#### 4. Meta-analysis of $d$ values using artifact distributions.

This program is used when (1) the meta-analysis is based on the effect size ( $d$  value; standardized difference between groups), and (2) information on the reliability of the measure of the dependent variable is *not* available in most of the primary studies. (This program is referred to in the text as D-VALUE1; see [Chapter 7](#).)

The “Type of Analyses” page of the program shows four types of meta-analysis (1 to 4, listed above), and the user selects the most appropriate analysis. If the user selects option 1 or 2, the subprograms (1A and 1B or 2A and 2B) will appear on subsequent pages (under the Analysis section) for further choice of the exact program to be used (depending on whether range restriction is direct or indirect). As noted earlier, the user can also indicate that there is no range restriction.

---

## Data Management

After selecting the appropriate type of analysis, the user is presented with the “Setting Up the Data” page. Here the user can opt to input (enter) new data, load existing (previously saved) data files, or import data files from Excel (when selecting this option, the user will be provided with an Excel template for entering data). In all programs, users can elect to code studies by potential moderators to make subsequent moderator subgroup analysis more convenient. This is done by following the instructions in the programs (described in Section 9, “Moderator Analysis”). The users can also conduct analysis to examine publication bias based on the cumulative meta-analysis method described in [Chapter 13](#) of the book.

### a. Entering Data

For meta-analysis type 1 (i.e., meta-analysis for correlations, with information on statistical and measurement artifacts available in most primary studies; VG6 programs), the user need only enter data into one general data file. This data file can be imported from Excel or created with the program.

The “Enter data from primary studies” page has a spreadsheet-like layout with six fields (spaces) so

that relevant information for each study (i.e., code for moderator analysis [explained later in Section 9, “Moderator Analysis”], correlation, sample size, reliability of variable  $X$  [independent variable reliability,  $R_{xx}$ ], reliability of variable  $Y$  [dependent variable reliability,  $R_{yy}$ ], and range restriction ratio [ $u$ ]) can be entered accordingly. Data are entered sequentially for each study. When corrections are needed, the user can click on the “Modification” buttons available in front of the data holder for each study. All the data fields must be filled. When there is no range restriction in a particular study, “1” should be entered in the range restriction ( $u$ ) cell. Similarly, when a variable is assumed to be perfectly measured (very rare case), “1” should be entered in the corresponding reliability cell. In situations where information on an artifact ( $R_{xx}$ ,  $R_{yy}$ , or  $u$ ) of a study is not available, the user can simply enter “99” into the corresponding cell. The program will then automatically use the mean of all the relevant artifact values provided in other studies to replace the missing value.

There are spaces to enter data for eight studies in each page. After entering the data for each page, the user clicks on the “Continue” button to proceed to the next page. The two buttons, “Back” and “Continue,” can be used to navigate through the pages to modify and/or enter data. The maximum number of studies is 1,000. After completing data entry, the user can press the “Done” button to exit to the previous page and start analysis (or choose other options, such as printing, saving, or modifying the data, as described later).

For meta-analysis type 2 (i.e., meta-analysis for correlations when information on statistical and measurement artifacts is *unavailable* in most primary studies; INTNL programs), the user enters data separately into several data files: The first data file consists of the code for moderators, correlations, and the corresponding sample sizes of the primary studies ( $r$  and  $N$ ). The second data file consists of the distribution of reliability coefficients of the independent variable ( $R_{xx}$  and freq of each). The third data file consists of the distribution of reliability coefficients of the dependent variable ( $R_{yy}$  and freq of each). And the fourth data file consists of the distribution of range restriction ( $u$  and freq of each  $u$ ). These data files can also be imported from Excel. If information is not available (or the artifact is not applicable) for an artifact distribution (e.g., when there is no range restriction), the program will assume that such artifacts have values fixed at 1.00 and automatically place 1.00s in the relevant data file(s). This means no correction will be made for these artifacts.

For meta-analysis type 3 (i.e., meta-analysis for effect sizes [ $d$  values], with information on statistical and measurement artifacts available in most primary studies; D-VALUE program), the procedures are similar to those of the type 1 meta-analysis described previously, except that the user enters only information on



dependent variable reliability ( $R_{yy}$ ); information on independent variable reliability ( $R_{xx}$ ) and range restriction ( $u$ ) is not required. This data file can be imported from Excel.

For meta-analysis type 4 (i.e., meta-analysis for effect sizes [ $d$  values] when information on statistical and measurement artifacts is *unavailable* in most primary studies; D-VALUE1 program), the procedures are similar to those of the type 2 meta-analysis described previously, except that the user enters only information on the distribution of dependent variable reliability ( $R_{yy}$  and freq); information on distributions of independent variable reliability ( $R_{xx}$ ) and range restriction ( $u$ ) is not required. Hence, there are only two data files rather than the four data files required for meta-analysis type 2. Again, these data files can be imported from Excel.

#### b. Saving Data

After completing the entry of the data (by clicking on the “Done” button), the user is taken back to the previous page where several options are presented: “Save,” “Print,” “Analysis,” and “Exit.” Selecting the “Save” option allows the user to save the data file he or she has just entered. The user will be asked to provide the name for the data set so that it can be easily retrieved when needed. The data will then be saved at the following location: C:\ Meta Analysis Programs \ Data \ “datasetname,” with  $i$  being the number corresponding to the type of meta-analysis.

Alternatively, the user can simply start analyzing the data. After the meta-analysis results are presented, the user is given another opportunity to save the current data set.

#### c. Loading Previously Saved Data

To load the previously saved data, select the “Load” option in the “Setting Up the Data” page. The user will be presented with names of all the previously saved data sets. He or she can select the appropriate data set to load into the program by clicking on the name of that data file.

#### d. Viewing/Modifying Saved Data

After loading/entering data, the user can view the data by selecting the “Entering/Modifying” option. Data will be presented in the spreadsheet-like layout. Modifications (corrections) can be made by clicking the icon in front of each individual study.

#### e. Printing Data

The user can print the current data set (i.e., the data that have just been entered, loaded, or imported from Excel) for easy reviewing by selecting the “Print” option.

#### f. Conducting Publication Bias Analysis

Selecting the link “Publication bias analysis” will open a new page (instructions will be given by the program), which enables the users to conduct publication analysis based on the cumulative meta-analysis method described in [Chapter 13](#) of the book.

---

## Analyzing Data

After entering/loading/modifying the data, the user can start analyzing the data by clicking on the “Analyzing the Data” button. For correlation-based meta-analysis (i.e., types 1 and 2; VG6 and INTNL programs), if the user has previously indicated there is range restriction, the user will next be asked to indicate the nature of range restriction existing in his or her data (i.e., direct or indirect, which means selecting between type 1A [VG6-D program] or 1B [VG6-I program] or selecting between 2A [INTNL-D program] or 2B [INTNL-I program]).

Both direct range restriction programs (VG6-D and INTNL-D) automatically assume that the independent variable reliabilities ( $R_{XX}$ ) are from the unrestricted samples, and the dependent variable reliabilities ( $R_{YY}$ ) are from the restricted samples. These assumptions agree with the nature of data available in research and practice. (See the text, [Chapters 3](#), [4](#), and [5](#), for a more detailed discussion.)

Both indirect range restriction programs (VG6-I and INTNL-I) require the user to specify whether (1) the independent variable reliabilities ( $R_{XX}$ ) are from the restricted or unrestricted samples and (2) whether the range restriction ratios are for true scores ( $uT$ ) or observed scores ( $uX$ ). For all cases, the program assumes that the dependent variable reliabilities are from the restricted samples. (See the text, [Chapters 3](#), [4](#), and [5](#), for a detailed discussion.)

When there is no range restriction, it does not matter which type of analysis (A or B) the user chooses; the programs will provide identical results.

The user will be asked to provide the title for the analysis (e.g., “Interviews and job performance—Meta-analysis 1”) and the name of the output file where results will be saved.

---

## Reporting Results

Results of the analyses are provided three different ways: (1) on screen (partial output), (2) on disk (at C:\Meta Analysis Programs\Output\“filename”, with “filename” being the name the user provided for the current analysis), and (3) print out (optional; which can be activated by clicking on the “Printer” option). As noted earlier, the printed output is present both in table form (some of the output) and in the more traditional output format (complete output). Due to space limitations, only partial output is presented on the screen. Complete output is saved to disk and printed with the print-out option.

A full listing and description of the output of each program is given later. The following are *some* of the items provided as output of the analysis:

1. Number of correlations (or  $d$  values) and total sample size.
2. Mean true score correlation (or mean corrected  $d$  value), the corresponding standard deviation (true score correlation  $SD\rho$  or true effect size  $SD\delta$ ), and the corresponding variances. These values are corrected for the biasing effects of all the artifacts considered in the meta-analysis. These values are estimates of mean construct-level relationships. Credibility intervals and confidence intervals are also provided.
3. Weighted mean observed correlation (or  $d$  values), observed variance and observed standard deviation, and variance and standard deviation corrected for sampling error only. This is the bare-bones meta-analysis output.
4. Sampling error variance, percentage of the observed variance due to sampling error variance, variance accounted for by all artifacts combined, and percentage of the observed variance due to all the artifacts combined.
5. The correlation between observed values and artifact effects is also presented. This is the square root of the proportion of variance accounted for by all artifacts combined.
6. For types 1 and 2 analyses (i.e., correlation-based meta-analyses), certain output is provided that is relevant to employment or educational selection research. The VG6 and INTNL programs provide the mean true validity and its standard deviation. Credibility intervals and confidence intervals are also provided. The true validity (also called operational validity) is the correlation between the predictor ( $X$ ) and the criterion ( $Y$ ) corrected for all the artifacts except for the attenuating effect of measurement error in the predictor  $X$ . This value represents the mean correlation of the predictor *measures* with the criterion of interest. (In contrast, the true score correlation represents the mean construct-level correlation between

the independent variable and the dependent variable.)

---

## Illustrating Examples

The programs include several data sets used as examples in [Chapters 3, 4, and 7](#). There are examples representing all six types of meta-analyses discussed previously. The user can practice doing analyses based on these data sets to familiarize himself or herself with the programs.

---

## Moderator Analysis

The programs allow users to conduct analyses separately for each moderator value/category. Results will be presented for each value/category of the moderator. To specify moderator analysis, the user can enter values of the moderator of interest at the field labeled “Moderator value” when entering the data. Values of the moderator should be entered using a number from 1 to  $k$  (with  $k$  being the number of categories of the moderator) for each study. If no moderator analysis is needed, the user can leave the field for “Moderator Value” blank.

---

## Extras (Auxiliary Programs)

There are three utility programs in the form of Microsoft Excel templates that aid in examining and processing data before entering data into the meta-analysis programs. The user must have Microsoft Excel to use these programs. During the process of installing the main (meta-analysis) programs, these three utility programs will be automatically copied onto your hard drive at the following location: “C: \Meta Analysis Programs\ Extras.” The first program (“Composite.xls”) combines correlations within a study; it computes the correlation between a composite (summed) independent or dependent variable and the other variable. This program also computes the reliability of the composite measure. (Both these procedures are described in [Chapter 10](#) of the text.) The second program (“Formula to compute biserial r.xls”) computes the biserial correlation from a point-biserial correlation provided in a primary study. This conversion should be carried out when a continuous (and

normally distributed) variable has been artificially dichotomized in a primary study (as described in [Chapters 4, 6 and 7](#) of the text). The third program allows users to create the forest plots to examine their data. As noted earlier, these programs can be accessed at the start page that users find when first starting the program packages.

---

## Full Description of Output of Individual Programs

The standard program output is divided into three sections: (1) Main Meta-Analysis Output, which presents results corrected for all artifacts; (2) Bare-Bones Meta-Analysis Output, which presents results corrected for sampling error only; and (3) Validity Generalization Output, which presents validity results relevant to tests and other procedures used in employment and educational selection. Section 3 is provided only for meta-analyses of correlations (i.e., type 1 meta-analysis [based on the VG6 programs] and type 2 meta-analyses [based on the INTNL programs]). It is not provided for meta-analyses of  $d$  values (type 3 meta-analysis [based on the D-VALUE program] and type 4 meta-analysis [based on the D-VALUE1 program]). The sections of program output always appear in the same order: The main output is always presented first, followed by the bare-bones output, followed by the validity generalization output (if applicable). The bare-bones output is identical for VG6 and INTNL program output. Hence, to avoid repetition, we present this first below.

### A. Bare-bones output for VG6 (type 1 meta-analysis) and INTNL (type 2 meta-analysis)

1. Sample size weighted mean observed correlation.
2. Variance of correlations after removing sampling error variance.
3. Standard deviation ( $SD$ ) of correlations after removing sampling error variance. (This is the square root of Item 2.)
4. Sample size weighted variance of observed correlations.
5. Sample size weighted  $SD$  of observed correlations. (This is the square root of Item 4.)
6. Variance due to sampling error variance.
7.  $SD$  predicted from sampling error alone. (This is the square root of Item 6.)
8. Percentage variance of observed correlations due to sampling error variance.
9. The correlation between observed values of  $r$  and their sampling errors. (This is the square root of the proportion of variance accounted for by

sampling error; see Item 8.)

B. Main output for VG6 programs (type 1 meta-analysis)

1. Number of correlations in the meta-analysis.
2. Total sample size. (Sum of study sample sizes.)
3. Mean true score correlation ( $\bar{\rho}$ ).
4. Variance of true score correlations ( $S_{\rho}^2$ ).
5. *SD* of true score correlations ( $SD_{\rho}$ ). (This is the square root of Item 4.) Note: For most purposes, the key output is Items 3 and 5.
6. Eighty percent credibility interval for true score correlation distribution (see [Chapter 5](#)).
7. Ninety-five percent confidence interval around the mean true score correlation (see [Chapter 5](#)).
8. Observed variance of the corrected correlations ( $S_{r_c}^2$ ). (Each *c* correlation is first corrected for measurement error and other artifacts; then the variance of these corrected correlations is computed. This is the variance of the corrected correlations before sampling error variance is removed. As described in [Chapter 3](#), the corrections for artifacts, while eliminating systematic downward biases, increase sampling error.)
9. *SD* of the corrected correlations  $SD_{r_c}$ . (This is the square root of Item 8.)
10. Variance in corrected correlations due to sampling error. (Note: This figure is larger than the variance in *uncorrected* [observed] correlations due to sampling error variance, which is reported in the bare-bones output section. This is because the artifact corrections, while removing systematic downward biases, increase sampling error variance. Note: Variance due to other artifacts is included in this variance because the effects of these artifacts have been previously corrected for.)
11. *SD* of corrected correlations predicted from sampling error. (This is the square root of Item 10.)
12. Percentage variance in corrected correlations due to sampling error and other artifacts.
13. The correlation between the corrected *rs* and their sampling errors. (This is the square root of the proportion of variance in the corrected correlations accounted for by all artifacts; see Item 12.)

C. Validity generalization output for VG6 programs (type 1 meta-analysis)

1. Mean true validity. (Same as mean true score correlation, except it is not corrected for the attenuating effects of measurement error in the independent variable; see [Chapter 3](#) of text.)
2. Variance of true validities.
3. *SD* of true validities. (This is the square root of Item 2.)
4. Eighty percent credibility interval for true validity distribution (see [Chapter 5](#)).
5. Ninety-five percent confidence interval around mean true validity (see [Chapter 5](#)).
6. Observed variance of the corrected validities. (Each validity is first corrected for measurement error in the dependent variable and for range restriction; then the variance of these corrected validities is computed. This is the variance of the corrected validities before sampling error is subtracted out. As described in [Chapter 3](#), the corrections for artifacts, while eliminating systematic downward biases, increase sampling error.)
7. *SD* of the corrected validities. (This is the square root of Item 6.)
8. Variance in corrected validities due to sampling error. (Note: This figure is larger than the variance in uncorrected [observed] validities due to sampling error variance, which is reported in the bare-bones output section. This is because the artifact corrections, while removing systematic downward biases, increase sampling error. Note: Variance due to other artifacts is included in this variance because the effects of these artifacts have previously been corrected for.)
9. *SD* of corrected validities predicted from sampling error and other artifacts. (This is the square root of Item 8.)
10. Percentage of variance accounted for by sampling error. (Note: Variance due to other artifacts is included in this because these artifacts have previously been corrected for.)
11. The correlation between corrected *r*s and artifact effects. (This is the square root of the proportion of variance in the corrected *r*s accounted for by artifacts; see Item 10.)

D. Main output for INTNL programs (type 2 meta-analysis)

1. Number of correlations in the meta-analysis.

2. Total sample size. (Sum of study sample sizes.)
  3. Mean true score correlation  $(\bar{\rho})$ .
  4. Variance of true score correlations  $(S_{\rho}^2)$ .
  5. *SD* of true score correlations ( $SD_{\rho}$ ). (This is the square root of Item 4.) Note: For most purposes, the key output is Items 3 and 5.
  6. Eighty percent credibility interval for true score correlation distribution (see [Chapter 5](#)).
  7. Ninety-five percent confidence interval around mean true score correlation (see [Chapter 5](#)).
  8. Variance in observed correlations due to all artifacts combined (see [Chapter 4](#)).
  9. *SD* of observed correlations predicted from all artifacts. (This is the square root of Item 8.)
  10. Variance of observed correlations after removal of variance due to all artifacts (residual variance [ $SD_{res}$ ]; see [Chapter 4](#)).
  11. Percentage variance of observed correlations due to all artifacts.
  12. Correlation between observed *rs* and artifact effects. (This is the square root of the proportion of variance in observed *rs* accounted for by artifacts; see Item 10.)
- E. Validity generalization output for INTNL programs (type 2 meta-analysis)
1. Mean true validity. (Same as mean true score correlation, except not corrected for the attenuating effects of measurement error in the independent variable.)
  2. Variance of true validities.
  3. *SD* of true validities. (This is the square root of Item 2.)
  4. Eighty percent credibility interval of true validity distribution (see [Chapter 5](#)).
  5. Ninety-five percent confidence interval around mean true validity (see [Chapter 5](#)).
  6. Variance of observed validities due to all artifacts combined (see [Chapter 4](#)).
  7. *SD* of observed validities predicted from all artifacts. (This is the square root of Item 6.)
  8. Variance in observed validities after removal of variance due to all artifacts



(residual variance [ $SD_{res}$ ]; see [Chapter 4](#)).

9. Percentage variance in observed validities due to all artifacts.
  10. Correlation between observed validities and artifact effects. (This is the square root of the proportion of variance in observed validities accounted for by artifacts; see Item 9.)
- F. Bare-bones output for D-VALUE and D-VALUE1 programs (meta-analysis types 3 and 4; bare-bones output is identical for these two types of meta-analyses)
1. Sample size weighted mean effect size (mean  $d$  value).
  2. Variance of  $d$  values after removing sampling error variance.
  3.  $SD$  of  $d$  values after removing sampling error variance. (This is the square root of Item 2.)
  4. Sample size weighted variance of observed  $d$  values.
  5. Sample size weighted standard deviation of observed  $d$  values. (This is the square root of Item 4.)
  6. Variance in observed  $d$  values due to sampling error variance.
  7.  $SD$  predicted from sampling error variance alone. (This is the square root of Item 6.)
  8. Percentage variance in observed  $d$  values due to sampling error variance.
  9. Correlation between observed  $d$  values and their sampling errors. (This is the square root of the proportion of variance accounted for by sampling error; see Item 8.)
- G. Main output for D-VALUE program (type 3 meta-analysis)
1. Number of effect sizes ( $d$  values) in the meta-analysis.
  2. Total sample size. (Sample sizes summed across studies.)
  3. Mean true effect size ( $\bar{\delta}$ )
  4. Variance of true effect sizes ( $S_{\delta}^2$ )
  5.  $SD$  of delta ( $SD_{\delta}$ ). (This is the square root of Item 4.) Note: For most purposes, the key output is Items 3 and 5.
  6. Eighty percent credibility interval for delta distribution (see [Chapter 8](#)).
  7. Ninety-five confidence interval around the mean delta (see [Chapter 8](#)).
  8. Observed variance of corrected  $d$  values ( $S_{d_c}^2$ ). (Each  $d$  value is corrected for measurement error in the dependent variable; then the variance of these corrected  $d$  values is computed. This is the variance of the corrected  $d$

values before sampling error variance is removed. As described in [Chapter 7](#), the correction for measurement error, while eliminating the systematic downward bias, increases sampling error variance.)

9. Observed  $SD$  of the corrected  $d$  values ( $SD_{d_c}$ ). (This is the square root of Item 8.)
10. Variance in corrected  $d$  values due to sampling error. (Note: This figure is larger than the variance in *uncorrected* [observed]  $d$  values due to sampling error variance, which is presented in the bare-bones output section. This is because the correction for measurement error, while removing the systematic downward biases, increases the sampling error variance. Note: Variance due to measurement error differences is included in this variance figure, because measurement error has previously been corrected for.)
11.  $SD$  of corrected  $d$  values predicted from sampling error variance. (This is the square root of Item 10.)
12. Percentage variance in corrected  $d$  values due to sampling error variance. (Note: Variance due to measurement error differences is included in this value because the effects of this artifact have previously been corrected for.)
13. Correlation between corrected  $d$  values and their sampling errors. (This is the square root of the proportion of variance accounted for by sampling error; see Item 12.)

#### H. Main output for D-VALUE1 program (type 4 meta-analysis)

1. Number of effect sizes ( $d$  values) in the meta-analysis.
2. Total sample size. (Sample sizes summed across studies.)
3. Mean true effect size ( $\bar{\delta}$ ).
4. Variance of true effect sizes ( $S_{\delta}^2$ ).
5.  $SD$  of delta ( $SD_{\delta}$ ). (This is the square root of Item 4.) Note: For most purposes, the key output is Items 3 and 5.
6. Eighty percent credibility interval for delta distribution (see [Chapter 8](#)).
7. Ninety-five percent confidence interval around mean delta (see [Chapter 8](#)).
8. Variance in observed  $d$  values due to sampling error and measurement error differences between studies (see [Chapter 7](#)).
9.  $SD$  of observed  $d$  values predicted from sampling error and measurement error differences. (This is the square root of Item 8.)

10. Variance in observed  $d$  values after removal of variance due to sampling error and between-study measurement error differences (residual variance  $[S_{\text{res}}^2]$ ; see [Chapter 7](#)).
11. Percentage variance in observed  $d$  values due to sampling error and differences in measurement error.
12. Correlation between observed  $d$  values and the combined effects of sampling error and differences in measurement error. (This is the square root of the proportion of variance due to all artifacts in Item 11.)

This program package can be ordered by contacting Huy Le, Department of Management, University of Nevada at Las Vegas. Electronic mail: [huyanhle@gmail.com](mailto:huyanhle@gmail.com). (Frank Schmidt can be contacted at: [frank-schmidt@uiowa.edu](mailto:frank-schmidt@uiowa.edu).) A website for this program is being prepared and will be available soon.

## References

- AERA-APA-NCME. (1985). *Standards for educational and psychological testing* (4th ed.). Washington, DC: American Educational Research Association.
- AERA-APA-NCME. (1999). *Standards for educational and psychological testing* (5th ed.). Washington, DC: American Educational Research Association.
- Aguinis, H. (2001). Estimation of sampling variance of correlations in meta-analysis. *Personnel Psychology*, *54*, 569–590.
- Aguinis, H., & Gottfredson, R. K. (2010). Best practice recommendations for estimating interaction effects using moderated multiple regression. *Journal of Organizational Behavior*, *31*, 776–786.
- 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. (2011). Debunking myths and urban legends about meta-analysis. *Organizational Research Methods*, *14*, 306–331.
- 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.
- Albright, L. E., Glennon, J. R., & Smith, W. J. (1963). *The use of psychological tests in industry*. Cleveland, OH: Howard Allen.
- Alexander, R. A., Carson, K. P., Alliger, G. M., & Carr, L. (1987). Correcting doubly truncated correlations: An improved approximation for correcting the bivariate normal correlation when truncation has occurred on both variables. *Educational and Psychological Measurement*, *47*, 309–315.
- Alexander, R. A., Carson, K. P., Alliger, G. M., & Cronshaw, S. F. (1989). Empirical distributions of range restricted SDx in validity studies. *Journal of Applied Psychology*, *74*, 253–258.
- Allen, M., Hunter, J. E., & Donahue, W. A. (1988). *Meta-analysis of self report data on the effectiveness of communication apprehension treatment techniques*. Unpublished manuscript, Department of Communication, Wake Forest University.

Allen, M., Hunter, J. E., & Donahue, W. A. (1989). Meta-analysis of self-report data on the effectiveness of public speaking anxiety treatment techniques. *Communication Education, 38*, 54–76.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*(2), 341–358.

Aloe, A. M., Becker, B. J., & Pigott, T. D. (2010). An alternative to *R*-squared for assessing linear models of effect sizes. *Research Synthesis Methods, 1*, 272–283.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

American Psychological Association. (2008). Reporting standards for research in psychology. *American Psychologist, 63*, 839–851.

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F., & Chalmers, T. C. (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. *Journal of the American Medical Association, 268*, 240–248.

Aronson, E., Ellsworth, P., Carlsmith, J., & Gonzales, M. (1990). *Methods of research in social psychology* (2nd ed.). New York: McGraw-Hill.

Arthur, W., Bennett, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*(2), 234–243.

Arvey, R. D., Cole, D. A., Hazucha, J., & Hartanto, F. (1985). Statistical power of training evaluation designs. *Personnel Psychology, 38*, 493–507.

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

Baker, R., & Jackson, D. (2008). A new approach to outliers in meta-analysis. *Health Care Management Science, 23*, 151–162.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*, 543–554.

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

388–399.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1983). Effects of coaching programs on achievement test performance. *Review of Educational Research*, 53, 571–585.

Banks, G. C., Kepes, S., & Banks, K. P. (2012). Publication bias: The antagonist of meta-analytic reviews and effective policy making. *Educational Evaluation and Policy Analysis*, 34, 259–277.

Banks, G. C., Kepes, S., & McDaniel, M. A. (2012). Publication bias: A call for improved meta-analytic practice in the organizational sciences. *International Journal of Selection and Assessment*, 20, 182–196.

Barnett, V., & Lewis, T. (1978). *Outliers in statistical data*. New York: John Wiley.

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

Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannan, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *Personnel Psychology*, 31, 233–241.

Beal, D. J., Corey, D. M., & Dunlap, W. P. (2002). On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in the meta-analysis of correlations. *Journal of Applied Psychology*, 87, 583–589.

Becker, B. J. (1987). Applying tests of combined significance in meta-analysis. *Psychological Bulletin*, 102, 164–171.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, 41, 257–278.

Becker, B. J. (1989, March). *Model-driven meta-analysis: Possibilities and limitations*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Becker, B. J. (1992). Models of science achievement: Forces affecting male and female performance in school science. In Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., et al. (Eds.), *Meta-analysis for explanation: A casebook* (pp. 209–282). New York: Russell Sage.

Becker, B. J. (1996). The generalizability of empirical research results. In Benbow, C. P., & Lubinski, D. (Eds.), *Intellectual talent: Psychological and social issues* (pp. 363–383). Baltimore: Johns Hopkins University Press.

Becker, B. J. (2005). Failsafe *N* or file drawer number. In Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. Chichester, UK: John Wiley.

- Becker, B. J. (2009). Model based meta-analysis. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (2nd ed., pp. 377–396). New York: Russell Sage.
- Becker, B. J., & Schram, C. M. (1994). Examining explanatory models through research synthesis. In Cooper, H., & Hedges, L. V. (Eds.), *The handbook of research synthesis* (pp. 357–382). New York: Russell Sage.
- Bedeian, A. G., Taylor, S. G., & Miller, A. N. (2010). Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning & Education*, 9, 715–725.
- Begg, C. B. (1994). Publication bias. In Cooper, H., & Hedges, L. V. (Eds.), *The handbook of research synthesis* (pp. 399–409). New York: Russell Sage.
- Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank order correlation for publication bias. *Biometrics*, 50, 1088–1101.
- Berlin, J. A., & Ghersi, D. (2005). Preventing publication bias: Registries and prospective meta-analyses. In Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.), *Publication bias is meta-analysis: Prevention, assessment, and adjustments* (pp. 35–48). West Sussex, UK: Wiley.
- Bettencourt, B. A., Talley, A., Benjamin, A. J., & Valentine, J. (2006). Personality and aggressive behavior under provoking and neutral conditions: A meta-analytic review. *Psychological Bulletin*, 132, 751–777.
- Bijmolt, T. H. A., & Pieters, R. G. M. (2001). Meta-analysis in marketing when studies contain multiple measurements. *Marketing Letters*, 12, 157–169.
- Billings, R., & Wroten, S. (1978). Use of path analysis in industrial/organizational psychology: Criticisms and suggestions. *Journal of Applied Psychology*, 63, 677–688.
- Bloom, B. S. (1964). *Stability and change in human characteristics*. New York: John Wiley.
- Bobko, P. (1983). An analysis of correlations corrected for attenuation and range restriction. *Journal of Applied Psychology*, 68, 584–589.
- Bobko, P., & Reick, A. (1980). Large sample estimators for standard errors of functions of correlation coefficients. *Applied Psychological Measurement*, 4, 385–398.
- Bobko, P., Roth, P. L., & Bobko, C. (2001). Correcting the effect size of  $d$  for range restriction and unreliability. *Organizational Research Methods*, 4, 46–61.
- Bobko, P., & Stone-Romero, E. F. (1998). Meta-analysis may be another useful research tool but it is not a panacea. In Ferris, G. R. (Ed.), *Research in personnel and human resources management* (Vol. 16, pp.

359–397). Greenwich, CT: JAI Press.

Bonett, D. G. (2007). Transforming odds ratios into correlations for meta-analytic research. *American Psychologist*, 62, 254–255.

Bonett, D. G. (2008). Meta-analytic interval estimation for bivariate correlations. *Psychological Methods*, 13, 173–189.

Bonett, D. G. (2009). Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychological Methods*, 14, 225–238.

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology*, 89(5), 901–910.

Borenstein, M. (1994). The case for confidence intervals in controlled clinical trials. *Controlled Clinical Trials*, 15, 411–428.

Borenstein, M. (2005). Software for publication bias. In Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 193–220). West Sussex, UK: Wiley.

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

Borman, G. D., & Grigg, J. A. (2009). Visual and narrative interpretation. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 497–520). New York: Russell Sage.

Bozarth, J. D., & Roberts, R. R. (1972). Signifying significant significance. *American Psychologist*, 27, 774–775.

Brannick, M. T. (2006, August). *Comparison of sample size and inverse variance weights for the effect size r*. Paper presented at the first annual meeting of the Society for Research Synthesis Methodology, Cambridge, UK.

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

Brennan, R. L. (1983). *Elements of generalizability theory*. Iowa City, IA: ACT Publications.

Brogden, H. E. (1968). *Restriction in range*. Unpublished manuscript, Department of Psychology, Purdue



University, Lafayette, IN.

Brown, S. H. (1981). Validity generalization and situational moderation in the life insurance industry. *Journal of Applied Psychology*, 66, 664–670.

Brozek, J., & Tiede, K. (1952). Reliable and questionable significance in a series of statistical tests. *Psychological Bulletin*, 49, 339–344.

Bryant, N. D., & Gokhale, S. (1972). Correcting correlations for restrictions in range due to selection on an unmeasured variable. *Educational and Psychological Measurement*, 32, 305–310.

Burke, M. J., & Day, R. (1986). A cumulative study of the effectiveness of management training. *Journal of Applied Psychology*, 71(2), 232–245.

Burke, M. J., & Landis, R. S. (2003). Methodological and conceptual challenges in conducting and interpreting meta-analyses. In Murphy, K. (Ed.), *Validity generalization: A critical review* (pp. 287–310). Mahwah, NJ: Lawrence Erlbaum.

Bushman, B. J., & Wang, M. D. (2009). Vote counting procedures in meta-analysis. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 207–220). New York: Russell Sage.

Callender, J. C. (1983, March). *Conducting validity generalization research based on correlations, regression slopes, and covariances*. Paper presented at the I/O and OB Graduate Student Convention, Chicago.

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

Callender, J. C., & Osburn, H. G. (1981). Testing the constancy of validity with computer generated sampling distributions of the multiplicative model variance estimate: Results for petroleum industry validation research. *Journal of Applied Psychology*, 66, 274–281.

Callender, J. C., & Osburn, H. G. (1988). Unbiased estimation of the sampling variance of correlations. *Journal of Applied Psychology*, 73, 312–315.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Campbell, J. P. (1990). The role of theory in industrial and organizational psychology. In Dunnette, M. D., & Hough, L. M. (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 39–73). Palo

Alto, CA: Consulting Psychologists Press.

Carey, B. (2011, November 2). Fraud case seen as red flag for psychology research. *New York Times*.

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

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology*, 84, 851–862.

Carlson, K. D., Scullen, S. E., Schmidt, F. L., Rothstein, H. R., & Erwin, F. W. (1999). Generalizable biographical data validity: Is multi-organizational development and keying necessary? *Personnel Psychology*, 52, 731–756.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review*, 48, 378–399.

Cattin, P. (1980). The estimation of the predictive power of a regression model. *Journal of Applied Psychology*, 65, 407–414.

Chan, A. W., & Altman, D. G. (2005). Outcome reporting bias in randomized trials on PubMed: Review of publications and survey of authors. *British Medical Journal*, 330, 753.

Chan, A. W., Hrobjartsson, A., Haahr, M. T., Gotzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *Journal of the American Medical Association*, 291, 2457–2465.

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

Chase, L. J., & Chase, R. B. (1976). Statistical power analysis of applied psychological research. *Journal of Applied Psychology*, 61, 234–237.

Chelimsky, E. (1994, October). *Use of meta-analysis in the General Accounting Office*. Paper presented at the Science and Public Policy Seminars, Federation of Behavioral, Psychological and Cognitive Sciences, Washington, DC.

Cheung, M. W. L. (2008). A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. *Psychological Methods*, 13, 182–202.

Cheung, M. W. L. (2009). Constructing approximate confidence intervals for parameters with structural

equation models. *Structural Equation Modeling*, 16, 267–294.

Cheung, M. W. L. (2010). Fixed-effects meta-analyses as multiple-group structural equation models. *Structural Equation Modeling*, 17, 481–509.

Cheung, M. W. L. (2012a). *MetaSEM: An R package for meta-analysis using structural equation modeling*. Manuscript under review.

Cheung, M. W. L. (2012b). *Three-level meta-analyses as structural equation models*. Manuscript under review.

Cheung, M. W. L. (in press). Multivariate meta-analysis as structural equation modeling. *Structural Equation Modeling*.

Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: a two-stage approach. *Psychological Methods*, 10, 40–64.

Cheung, S. F., & Change, D. K. (2004). Dependent effects sizes in meta-analysis: Incorporating the degree of interdependence. *Journal of Applied Psychology*, 89, 780–791.

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19, 3127–3131.

Clarke, M. (2009). Reporting format. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 521–534). New York: Russell Sage.

Coffman, D. L., & MacCallum, R. C. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research*, 40, 235–259.

Coggin, T. D., & Hunter, J. E. (1987). A meta-analysis of pricing of “risk” factors in APT. *Journal of Portfolio Management*, 14, 35–38.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145–153.

Cohen, J. (1977). *Statistical power analysis for the behavior sciences* (Rev. ed.). New York: Academic Press.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, 7, 249–253.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

- Cohen, J. (1990). Things I learned (so far). *American Psychologist*, 45, 1304–1312.
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, 1, 98–101.
- Cohen, J. (1994). The earth is round ( $P < .05$ ). *American Psychologist*, 49, 997–1003.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Coleman, J. S. (1966). *Equality of educational opportunity*. Washington, DC: Government Printing Office.
- Collins, D. B., & Holton, E. F. (2004). The effectiveness of managerial leadership development programs: a meta-analysis of studies from 1982 to 2001. *Human Resource Development Quarterly*, 15, (2), 217–248.
- Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can individual differences shed light on the construct meaning of assessment centers? *International Journal of Selection and Assessment*, 11, 17–29.
- 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.
- Cook, T., & Campbell, D. T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In Dunnette, M. (Ed.), *Handbook of industrial and organizational psychology* (pp. 223–236). Chicago: Rand McNally.
- Cook, T., & Campbell, D. T. (1979). *Quasi-experiments and true experimentation: Design and analysis for field settings*. Chicago: Rand McNally.
- Cook, T. D., Cooper, H., Cordray, D. S., Hartmann, H., Hedges, L. V., Light, R. J., et al. (1992). *Meta-analysis for explanation: A casebook*. New York: Russell Sage.
- Cooper, H. (1997). Some finer points in meta-analysis. In Hunt, M. (Ed.), *How science takes stock: The story of meta-analysis* (pp. 169–181). New York: Russell Sage.
- Cooper, H. (1998). *Synthesizing research: A guide for literature reviews*. Thousand Oaks, CA: Sage.
- Cooper, H. (2003). Editorial. *Psychological Bulletin*, 129, 3–9.
- Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Los Angeles: Sage.

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

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin*, 87, 442–449.

Copas, J. B. (1999). What works? Selectivity models and meta-analysis. *Journal of the Royal Statistical Society, Series A*, 162, 95–109.

Copas, J. B., & Shi, J. Q. (2001). A sensitivity analysis for publication bias in systematic reviews. *Statistical Methods in Medical Research*, 10, 251–265.

Cordray, D. S., & Morphy, P. (2009). Research synthesis and public policy. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 473–494). New York: Russell Sage.

Coursol, A., & Wagner, E. E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology*, 17, 136–137.

Coward, W. M., & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology*, 75, 297–300.

Cronbach, L. J. (1947). Test “reliability”: Its meaning and determination. *Psychometrika*, 12, 1–16.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology revisited. *American Psychologist*, 30, 116–127.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Cureton, E. E. (1936). On certain estimated correlation functions and their standard errors. *Journal of Experimental Education*, 4, 252–264.

Cuts raise new social science query: Does anyone appreciate social science? (1981, March 27). *Wall Street Journal*, p. 54.

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer

problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology*, 65, 221–249.

Dean, M. A., Roth, P. L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: a meta-analysis. *Journal of Applied Psychology*, 93(3), 685–691.

DeGeest, D. S., & Schmidt, F. L. (2011). The impact of research synthesis methods on industrial-organizational psychology: The road from pessimism to optimism about cumulative knowledge. *Research Synthesis Methods*, 1, 185–197.

Dickersin, K. (1994). Research registers. In Cooper, H., & Hedges, L. V. (Eds.), *The handbook of research synthesis* (pp. 71–84). New York: Russell Sage.

Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. In Rothstein, H., Sutton, A. J., & Borenstein, M. (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 11–34). Chichester, UK: Wiley.

Dickersin, K., Min, Y., & Meinert, C. (1992). Factors influencing the publication of research results: Follow-up of applications submitted to two institutional review boards. *Journal of the American Medical Association*, 267, 374–378.

Dieckmann, N. F., Malle, B. F., & Bodner, T. E. (2009). An empirical assessment of meta-analytic practice. *Review of General Psychology*, 13, 101–115.

Doucouliagos, C., & Stanley, T. D. (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, 47, 406–428.

Doucouliagos, C., & Stanley, T. D. (2011). Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys*, 10, 1–29.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1, 170–177.

Dunnette, M. D., Houston, J. S., Hough, L. M., Touquam, J., Lamnstein, S., King, K., et al. (1982). *Development and validation of an industry-wide electric power plant operator selection system*. Minneapolis, MN: Personnel Decisions Research Institute.

Duval, S. (2005). The trim and fill method. In Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 127–144). New York: John Wiley.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 276–284.

Dye, D. (1982). *Validity generalization analysis for data from 16 studies participating in a consortium study*. Unpublished manuscript, Department of Psychology, George Washington University, Washington, DC.

Dye, D., Reck, M., & Murphy, M. A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment*, 1, 153–157.

Eagly, A. H. (1978). Sex differences in influenceability. *Psychological Bulletin*, 85, 86–116.

Eagly, A. H., Johannsen-Schmidt, M. C., & van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin*, 129, 569–591.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin*, 117, 125–145.

Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D. R. (1991). Publication bias in clinical research. *Lancet*, 337, 867–872.

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634.

Emerson, G. B., Warne, W. J., Wolf, F. M., Heckman, J. D., Brand, R. A., & Leopold, S. S. (2010). Testing for the presence of positive-outcome bias in peer review: A randomized controlled trial. *Archives of Internal Medicine*, 170, 1934–1939.

Erlenmeyer-Kimling, L., & Jarvik, L. F. (1963). Genetics and intelligence: A review. *Science*, 142, 1477–1479.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4, 1–11.

Fiedler, K. (2011). Voodoo correlations are everywhere—not just in neuroscience. *Perspectives on Psychological Science*, 6, 163–171.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods*, 6, 161–180.

Field, A. P. (2005). Is the meta-analysis of correlations accurate when population correlations vary. *Psychological Methods*, 10, 444–467.



- Fisher, R. A. (1932). *Statistical methods for research workers* (4th ed.). London: Oliver & Boyd.
- Fisher, R. A. (1935). *The design of experiments*. London: Oliver & Boyd.
- Fisher, R. A. (1938). *Statistical methods for research workers* (7th ed.). London: Oliver & Boyd.
- Forsyth, R. A., & Feldt, L. S. (1969). An investigation of empirical sampling distributions of correlation coefficients corrected for attenuation. *Educational and Psychological Measurement*, 29, 61–71.
- Fountoulakis, K. N., Conda, X., Vieta, E., & Schmidt, F. L. (2009). Treatment of psychotic symptoms in bipolar disorder with aripiprazole monotherapy. *Annals of General Psychiatry*, 8, 27. Available at [www.annals-general-psychiatry.com/contents/8/1/27](http://www.annals-general-psychiatry.com/contents/8/1/27)
- Francis, G. (2012a). The psychology of replication and replication in psychology. *Perspectives on Psychological Science*, 7, 585–594.
- Francis, G. (2012b). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, 19, 151–156.
- Francis, G. (2013). Publication bias in “Red, Rank, and Romance in Women Viewing Men” by Elliot et al. (2010). *Journal of Experimental Psychology: General*, 142, 292–296.
- Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, 138, 96–321.
- Gardner, S., Frantz, R. A., & Schmidt, F. L. (1999). The effect of electrical stimulation on chronic wound healing: A meta-analysis. *Nursing Research*, 7, 495–403.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493–511.
- Gendreau, P., & Smith, P. (2007). Influencing the people who count: Some perspectives on reporting of meta-analysis results for prediction and treatment of outcomes with offenders. *Criminal Justice and Behavior*, 34, 1536–1559.
- Gergen, K. J. (1982). *Toward transformation in social knowledge*. New York: Springer-Verlag.
- 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.
- Ghiselli, E. E. (1949). The validity of commonly employed occupational tests. *University of California*



*Publications in Psychology*, 5, 253–288.

Ghiselli, E. E. (1955). The measurement of occupational aptitude. *University of California Publications in Psychology*, 8, 101–216.

Ghiselli, E. E. (1966). *The validity of occupational aptitude tests*. New York: John Wiley.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, 26, 461–477.

Gigerenzer, G. (2007). Helping physicians understand screening tests will improve health care. *Association for Psychological Science Observer*, 20, 37–38.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest*, 8, 53–96.

Glass, G. V. (1972). The wisdom of scientific inquiry on education. *Journal of Research in Science Teaching*, 9, 3–18.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 5, 3–8.

Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351–379.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying fixed effects analysis of variance and covariance. *Review of Educational Research*, 42, 237–288.

Gleser, L. J., & Olkin, O. (2009). Stochastically dependent effect sizes. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 357–376). New York: Russell Sage.

Gottfredson, L. S. (1985). Education as a valid but fallible signal of worker quality. *Research in Sociology of Education and Socialization*, 5, 123–169.

Green, B. F., & Hall, J. A. (1984). Quantitative methods for literature reviews. *Annual Review of Psychology*, 35, 37–53.

Greenhouse, J. B., & Iyengar, S. (1994). Sensitivity analysis and diagnostics. In Hedges, L. V., & Cooper, H. (Eds.), *Handbook of research synthesis* (pp. 383–398). New York: Russell Sage Foundation.

- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82, 1–20.
- Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research* (2nd ed.). New York: Routledge.
- Gross, A. L., & McGanney, M. L. (1987). The range restriction problem and nonignorable selection processes. *Journal of Applied Psychology*, 72, 604–610.
- Grubbs, F. E. (1969). Procedures for detecting outliers. *Technometrics*, 11, 1–21.
- Gulliksen, H. (1986). The increasing importance of mathematics in psychological research (Part 3). *The Score*, 9, 1–5.
- Guttman, L. (1985). The illogic of statistical inference for cumulative science. *Applied Stochastic Models and Data Analysis*, 1, 3–10.
- Guzzo, R. A., Jackson, S. E., & Katzell, R. A. (1986). Meta-analysis analysis. In Cummings, L. L., & Staw, B. M. (Eds.), *Research in organizational behavior* (Vol. 9). Greenwich, CT: JAI Press.
- Hackman, J. R., & Oldham, G. R. (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology*, 60, 159–170.
- Haddock, C., Rindskopf, D., & Shadish, W. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods*, 3, 339–353.
- Hafdahl, A. R. (2009). Improved Fisher's  $z$  estimators for univariate random-effects meta-analysis of Correlations. *British Journal of Mathematical and Statistical Psychology*, 62, 233–261.
- Hafdahl, A. R. (2010). Random-effects meta-analysis of correlations: Evaluation of mean estimates. *British Journal of Mathematical and Statistical Psychology*, 63, 227–254.
- Hafdahl, A. R. (2012). Article alerts: Items from 2011. *Research Synthesis Methods*, 3, 325–331.
- Hafdahl, A. R., & Williams, M. A. (2009). Meta-analysis of correlations revisited: Attempted replication and extension of Field's (2001) simulation studies. *Psychological Methods*, 14, 24–42.
- Hall, S. M., & Brannick, M. T. (2002). Comparison of two random effects methods of meta-analysis. *Journal of Applied Psychology*, 87, 377–389.
- Halvorsen, K. T. (1994). The reporting format. In Cooper, H., & Hedges, L. V. (Eds.), *Handbook of research synthesis* (pp. 425–438). New York: Russell Sage.

Hamilton, M. A., & Hunter, J. E. (1987, August). *Two accounts of language intensity effects*. Paper presented at the International Communication Association Convention, New Orleans, LA.

Harmon, C., Oosterbeek, H., & Walker, I. (2000). *The returns to education: A review of evidence, issues and deficiencies in the literature* (Discussion Paper No. 5). London: Center for the Economics of Educations (CEE), London School of Economics.

Harter, J. K., Schmidt, F. L., Asplund, J., & Killham, E. A. (2010). Casual impact of employee work perceptions on the bottom line of organizations. *Perspectives on Psychological Science*, 5, 378–389.

Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business unit level relationships between employee satisfaction/engagement and business outcomes: A meta-analysis. *Journal of Applied Psychology*, 87, 268–279.

Hartigan, J. A., & Wigdor, A. K. (Eds.). (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery*. Washington, DC: National Academies Press.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107–128.

Hedges, L. V. (1982a). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92, 490–499.

Hedges, L. V. (1982b). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics*, 7, 119–137.

Hedges, L. V. (1982c). Fitting continuous models to effect size data. *Journal of Educational Statistics*, 7, 245–270.

Hedges, L. V. (1983a). Combining independent estimators in research synthesis. *British Journal of Mathematical and Statistical Psychology*, 36(1), 123–131.

Hedges, L. V. (1983b). A random effects model for effect sizes. *Psychological Bulletin*, 93, 388–395.

Hedges, L. V. (1984). Estimation of effect size under non-random sampling: The effects of censoring studies yielding statistically mean differences. *Journal of Educational Statistics*, 9, 61–85.

Hedges, L. V. (1987). How hard is hard science, how soft is soft science: The empirical cumulativeness of research. *American Psychologist*, 42, 443–455.

Hedges, L. V. (1989). An unbiased correction for sampling error in validity generalization studies. *Journal of*

*Applied Psychology*, 74, 469–477.

Hedges, L. V. (1992b). Modeling publication selection effects in meta-analysis. *Statistical Science*, 7, 246–255.

Hedges, L. V. (1995, February 14). *Letter to Professor Herman Aguinis explaining sampling error variance of corrected correlations*.

Hedges, L. V. (2009a). Effect sizes in nested designs. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (2nd ed., pp. 337–356). New York: Russell Sage.

Hedges, L. V. (2009b). Statistical considerations. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (2nd ed., pp. 37–48). New York: Russell Sage.

Hedges, L. V., & Olkin, I. (1980). Vote counting methods in research synthesis. *Psychological Bulletin*, 88, 359–369.

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

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods*, 6, 203–217.

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9, 426–425.

Hedges, L. V., & Stock, W. (1983). The effects of class size: An examination of rival hypotheses. *American Educational Research Journal*, 20, 63–85.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010a). Erratum: Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1, 164–165.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010b). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1, 39–65.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3, 486–504.

Hedges, L. V., & Vevea, J. (2005). Selection method approaches. In Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. Chichester, UK: John Wiley.

Heinsman, D. T., & Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate the answers from randomized experiments? *Psychological Methods*, 1, 154–169.

Higgins, J. P. T., & Thompson, S. G. (2001, October). *Presenting random effects meta-analyses: Where are we going wrong?* Paper presented at the 9th International Cochrane Colloquium, Lyon, France.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal*, 327, 557–560.

Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society*, 172(Pt. 1), 137–159.

Hill, T. E. (1980, September). *Development of a clerical program in Sears*. In V. J. Benz (Chair), *Methodological implications of large scale validity studies of clerical occupations*. Symposium conducted at the meeting of the American Psychological Association, Montreal, Canada.

Hirsh, H. R., Northrop, L. C., & Schmidt, F. L. (1986). Validity generalization results for law enforcement occupations. *Personnel Psychology*, 39, 399–420.

Hoffert, S. P. (1997). Meta-analysis is gaining status in science and policymaking. *The Scientist*, 11(18), 1–6.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills: Sage.

Hotelling, H. (1953). New light on the correlation coefficient and its transforms. *Journal of the Royal Statistical Society, B*, 15, 193–225.

Hoyt, W. T. (2000). Rater bias in psychological research: When it is a problem and what we can do about it? *Psychological Methods*, 5, 64–86.

Huber, P. J. (1980). *Robust statistics*. New York: John Wiley.

Huffcutt, A. I., & Arthur, W. A. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, 80, 327–334.

Huffcutt, A. I., Arthur, W. A., & Bennett, W. (1993). Conducting meta-analysis using the Proc Means procedure in SAS. *Educational and Psychological Measurement*, 53, 119–131.

Hunt, M. (1997). *How science takes stock*. New York: Russell Sage.

Hunter, J. E. (1980). Factor analysis. In Monge, P. (Ed.), *Multivariate techniques in human communication*

*research*. New York: Academic Press.

Hunter, J. E. (1983a). A causal analysis of cognitive ability, job knowledge, job performance, and supervisory ratings. In Landy, F., Zedeck, S., & Cleveland, J. (Eds.), *Performance measurement and theory* (pp. 257–266). Hillsdale, NJ: Lawrence Erlbaum.

Hunter, J. E. (1983b). *Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the general aptitude test battery (GATB)* (Test Research Rep. No. 45). Washington, DC: U.S. Department of Labor, U.S. Employment Service.

Hunter, J. E. (1986, November). *Multiple dependent variables in experimental design*. Monograph presented at a workshop at the University of Iowa, Iowa City.

Hunter, J. E. (1987). Multiple dependent variables in program evaluation. In Mark, M. M., & Shotland, R. L. (Eds.), *Multiple methods in program evaluation*. San Francisco: Jossey-Bass.

Hunter, J. E. (1988). *A path analytic approach to analysis of covariance*. Unpublished manuscript, Department of Psychology, Michigan State University, East Lansing.

Hunter, J. E. (1995). *PACKAGE: Software for data analysis in the social sciences*. Unpublished suite of computer programs. (Available from Frank Schmidt, University of Iowa.)

Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science*, 8, 3–7.

Hunter, J. E., & Gerbing, D. W. (1982). Unidimensional measurement, second order factor analysis and causal models. In Staw, B. M., & Cummings, L. L. (Eds.), *Research in organizational behavior* (Vol. 4). Greenwich, CT: JAI Press.

Hunter, J. E., & Hirsch, H. R. (1987). Applications of meta-analysis. In Cooper, C. L., & Robertson, I. T. (Eds.), *International review of industrial and organizational psychology 1987*. London: Wiley.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin*, 96, 72–98.

Hunter, J. E., & Schmidt, F. L. (1977). A critical analysis of the statistical and ethical implications of various definitions of test fairness. *Psychological Bulletin*, 83, 1053–1071.

Hunter, J. E., & Schmidt, F. L. (1987a). *Error in the meta-analysis of correlations: The mean correlation*. Unpublished manuscript, Department of Psychology, Michigan State University, East Lansing.

Hunter, J. E., & Schmidt, F. L. (1987b). *Error in the meta-analysis of correlations: The standard deviation*.

Unpublished manuscript, Department of Psychology, Michigan State University, East Lansing.

Hunter, J. E., & Schmidt, F. L. (1990a). Dichotomizing continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, 75, 334–349.

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

Hunter, J. E., & Schmidt, F. L. (1994). The estimation of sampling error variance in meta-analysis of correlations: The homogeneous case. *Journal of Applied Psychology*, 79, 171–177.

Hunter, J. E., & Schmidt, F. L. (1996). Cumulative research knowledge and social policy formulation: The critical role of meta-analysis. *Psychology, Public Policy, and Law*, 2, 324–347.

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

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

Hunter, J. E., Schmidt, F. L., & Coggin, T. D. (1996). *Meta-analysis of correlations: Bias in the correlation coefficient and the Fisher z transformation*. Unpublished manuscript, University of Iowa, Iowa City.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, 31, 215–232.

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.

Ioannidis, J. P. (2005a). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association*, 294, 218–226.

Ioannidis, J. P. (2005b). Why most published research findings are false. *PLoS Medicine*, 2(8), 696–701.

Ioannidis, J. P., & Doucouliagos, C. (2013). What's to know about the credibility of empirical economics? *Journal of Economic Surveys*, 13, 1–8.

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical*



*Trials*, 4, 245–253.

Iyengar, S., & Greenhouse, J. (1988). Selection models and the file drawer problem. *Statistical Science*, 3, 109–135.

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

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524–532.

Johnson, B. T. (1989). *D-Stat: Software for the meta-analytic review of research literatures*. Hillsdale, NJ: Lawrence Erlbaum.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability—with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, 86, 80–92.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, 87(4), 765–780.

Judge, T. A., Colbert, A.E., & Ilies, R. (2004). Intelligence and leadership: a quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, 89(3), 542–552.

Judge, T. A., Piccolo, R. F., & Kosalka, T. (2009). The bright and dark sides of leader traits: A review theoretical extension of the leader trait paradigm. *Leadership Quarterly*, 20(6), 855–875.

Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). The job satisfaction-job performance relationship: A qualitative and quantitative review. *Psychological Bulletin*, 127, 376–401.

Kelly, T. L. (1947). *Fundamentals of statistics*. Cambridge, MA: Harvard University Press.

Kemery, E. R., Dunlap, W. P., & Griffeth, R. W. (1988). Correction for unequal proportions in point biserial correlations. *Journal of Applied Psychology*, 73, 688–691.

Kemery, E. R., Mossholder, K. W., & Roth, L. (1987). The power of the Schmidt and Hunter additive model of validity generalization. *Journal of Applied Psychology*, 72, 30–37.

Kepes, S., Banks, G. C., & Oh, I.-S. (in press). Avoiding bias in publication bias research: The value of “null”



findings. *Journal of Business and Psychology*.

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

Kepes, S., & McDaniel, M. A. (2013). How trustworthy is the scientific literature in I-O psychology? *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 6(3), 252–268.

Killeen, P. R. (2005a). An alternative to null hypothesis significance tests. *Psychological Science*, 16, 345–353.

Killeen, P. R. (2005b). Replicability, confidence, and priors. *Psychological Science*, 16, 1009–1012.

King, L. M., Hunter, J. E., & Schmidt, F. L. (1980). Halo in multidimensional forced choice performance evaluation scale. *Journal of Applied Psychology*, 65, 507–516.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences*. New York: Brooks/Cole.

Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement*, 61, 213–218.

Kirkpatrick, D. L. (2000). Evaluating training programs: the four levels. In Piskurich, G. M., Beckschi, P., & Hall, B. (Eds.), *The ASTD handbook of training design and delivery* (pp. 133–146). New York: McGraw-Hill.

Kisamore, J. L. (2003). *Validity generalization and transportability: An investigation of distributional assumptions of random-effects meta-analytic methods*. Unpublished doctoral dissertation, Department of Psychology, University of South Florida, Tampa.

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

Kline, R. B. (2004). *Beyond significant testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychology Association.

Kotov, R., Gamez, W., Schmidt, F. L., & Watson, D. (2010). Linking “Big” personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, 136, 768–821.

Krakovsky, M. (2004). Register or perish. *Scientific American*, 291, 18–20.

Kulik, J. A., & Bangert-Drowns, R. L. (1983-1984). Effectiveness of technology in precollege mathematics and science teaching. *Journal of Educational Technology Systems*, 12, 137–158.

- Kulinskaya, E., Morgenthaler, S., & Staudte, R. G. (2010). Combining the evidence using stable weights. *Research Synthesis Methods, 1*, 284–296.
- Laczo, R. M., Sackett, P. R., Bobko, P., & Cortina, J. M. (2005). A comment on sampling error in  $d$  with unequal  $N$ s: Avoiding potential errors in meta-analytic and primary research. *Journal of Applied Psychology, 90*, 758–764.
- Landis, R. S. (2013). Successfully combining meta-analysis and structural equation modeling: Recommendations and strategies. *Journal of Business and Psychology, 28*, 251–261.
- Landman, J. T., & Dawes, R. M. (1982). Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny. *American Psychologist, 37*, 504–516.
- Law, K. S. (1995). The use of Fisher's  $Z$  in Schmidt-Hunter type meta-analysis. *Journal of Educational and Behavioral Statistics, 20*, 287–306.
- Law, K. S., Schmidt, F. L., & Hunter, J. E. (1994a). Nonlinearity of range corrections in meta-analysis: A test of an improved procedure. *Journal of Applied Psychology, 79*, 425–438.
- Law, K. S., Schmidt, F. L., & Hunter, J. E. (1994b). A test of two refinements in meta-analysis procedures. *Journal of Applied Psychology, 79*, 978–986.
- Lawshe, C. H. (1948). *Principles of personnel selection*. New York: McGraw-Hill.
- Le, H. (2003). *Correcting for indirect range restriction in meta-analysis: Testing a new meta-analysis method*. Unpublished doctoral dissertation, University of Iowa, Iowa City.
- Le, H., & Schmidt, F. L. (2006). Correcting for indirect range restriction in meta-analysis: Testing a new meta-analysis procedure. *Psychological Methods, 11*, 416–438.
- Le, H., Schmidt, F. L., Harter, J. K., & Lauver, K. (2010). The problem of empirical redundancy of constructs in organizational research: An empirical investigation. *Organizational Behavior and Human Decision Processes, 112*, 112–123.
- Le, H., Schmidt, F. L., & Oh, I.-S. (2013). *Correction for range restriction in meta-analysis revisited: Improvements and implications for organizational research*. Manuscript under review.
- Le, H., Schmidt, F. L., & Putka, D. J. (2009). The multi-faceted nature of measurement error and its implications for measurement error corrections. *Organizational Research Methods, 12*, 165–200.
- Lehrer, J. (2011, December 13). The truth wears off: Is there something wrong with the scientific method?

*New Yorker Magazine*, pp. 52–56.

Lent, R. H., Auerbach, H. A., & Levin, L. S. (1971a). Predictors, criteria and significant results. *Personnel Psychology*, 24, 519–533.

Lent, R. H., Auerbach, H. A., & Levin, L. S. (1971b). Research design and validity assessment. *Personnel Psychology*, 24, 247–274.

Leone, F. C., & Nelson, L. S. (1966). Sampling distributions of variance components: I. Empirical studies of balanced nested designs. *Technometrics*, 8, 457–468.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, 61(2), 273–307.

Levine, J. M., Kramer, G. G., & Levine, E. N. (1975). Effects of alcohol on human performance: An integration of research findings based on an abilities classification. *Journal of Applied Psychology*, 60, 285–293.

Levine, J. M., Romashko, T., & Fleishman, E. A. (1973). Evaluation of an abilities classification system for integration and generalizing human performance research findings: An application to vigilance tasks. *Journal of Applied Psychology*, 58, 149–157.

Li, J. C., Chan, W., & Cui, Y. (2010). Bootstrap standard error and confidence intervals for correlations corrected for indirect range restriction. *British Journal of Mathematical and Statistical Psychology*, 64, 367–387.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Light, R. J., Singer, J. D., & Willett, J. B. (1994). The visual presentation and interpretation of meta-analyses. In Cooper, H., & Hedges, L. V. (Eds.), *The handbook of research synthesis* (pp. 439–453). New York: Russell Sage.

Light, R. J., & Smith, P. V. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review*, 41, 429–471.

Linn, R. L., Harnisch, D. L., & Dunbar, S. B. (1981a). Corrections for range restriction: An empirical investigation of conditions resulting in conservative corrections. *Journal of Applied Psychology*, 66, 655–663.

Linn, R. L., Harnisch, D. L., & Dunbar, S. B. (1981b). Validity generalization and situational specificity: An

analysis of the prediction of first year grades in law school. *Applied Psychological Measurement*, 5, 281–289.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181–1209.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Lockhart, R. S. (1998). *Statistics and data analysis for the behavioral sciences*. New York: W. H. Freeman.

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. New York: Knopf.

Lumley, T. (2009). *Rmeta meta-analysis. R package version 2.16*. <http://CRAN.R-project.org/package=rmeta>

Mabe, P. A. III, & West, S. G. (1982). Validity of self evaluations of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67, 280–296.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19–40.

MacMahon, S., Peto, R., Cutler, J., Collins, R., Sorlie, P., Neaton, J., et al. (1990). Blood pressure, stroke, and coronary heart disease: Part 1: Prolonged differences in blood pressure: Prospective observational studies corrected for the regression dilution bias. *Lancet*, 335, 763–774.

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

Maloley, v. et. al. *Department of National Revenue*, Canadian Civil Service Appeals Board, Ottawa (1986, February).

Mann, C. (1990, August 3). Meta-analysis in the breech. *Science*, 249, 476–480.

Mansfield, R. S., & Busse, T. V. (1977). Meta-analysis of research: A rejoinder to Glass. *Educational Researcher*, 6, 3.

Marin-Martinez, F., & Sanchez-Meca, J. (2010). Weighting by inverse variance or by sample size in random effects meta-analysis. *Educational and Psychological Measurement*, 70, 56–73.

Martinussen, M., & Bjornstad, J. F. (1999). Meta-analysis calculations based on independent and nonindependent cases. *Educational and Psychological Measurement*, 59, 928–950.

Matt, G. E., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (2nd ed., pp. 537–560). New

York: Russell Sage.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, 9, 147–163.

McDaniel, M. A. (1986a). Computer programs for calculating meta-analysis statistics. *Educational and Psychological Measurement*, 64, 175–177.

McDaniel, M. A. (1986b). *MAME: Meta-analysis made easy. Computer program and manual. Ver. 2.1.* Bethesda, MD: Author.

McDaniel, M. A. (1990, April 14). *Cumulative meta-analysis as a publication bias method*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.

McDaniel, M. A., & Nguyen, N. T. (2002, December 5). *A meta-analysis of the relationship between in vivo brain volume and intelligence*. Paper presented at the Third Annual Conference of the International Society for Intelligence Research, Nashville, TN.

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

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988a). Job experience correlates of job performance. *Journal of Applied Psychology*, 73, 327–330.

McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988b). A meta-analysis of the validity of training and experience ratings in personnel selection. *Personnel Psychology*, 41, 283–314.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, 79, 599–616.

McNatt, D. B. (2000). Ancient Pygmalion joins contemporary management: A meta-analysis of the result. *Journal of Applied Psychology*, 85, 314–322.

McNemar, Q. (1960). At random: Sense and nonsense. *American Psychologist*, 15, 295–300.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806–834.

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

- Mendoza, J. L., & Reinhardt, R. N. (1991). Validity generalization procedures using sample-based estimates: A comparison of six procedures. *Psychological Bulletin*, 110, 596–610.
- Mendoza, J. L., Stafford, K. L., & Stauffer, J. M. (2000). Large sample confidence intervals for validity and reliability coefficients. *Psychological Methods*, 5, 356–369.
- Miller, N., Lee, S., & Carlson, M. (1991). The validity of inferential judgments when used in theory-testing meta-analysis. *Personality and Social Psychology Bulletin*, 17, 335–343.
- Millsap, R. (1988). Sampling variance in attenuated correlation coefficients: A Monte Carlo study. *Journal of Applied Psychology*, 73, 316–319.
- Millsap, R. (1989). The sampling variance in the correlation under range restriction: A Monte Carlo study. *Journal of Applied Psychology*, 74, 456–461.
- Moher, D., & Olkin, I. (1995). Meta-analysis of randomized controlled trials: A concern for standards. *Journal of the American Medical Association*, 274, 1962–1964.
- Morris, S. B. (2007). [Review of the book *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.).] *Organizational Research Methods*, 11, 184–187.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11, 364–386.
- Morris, S. B., & DeShon, R. P. (1997). Correcting effect sizes computed from factorial analysis of variance for use in meta-analysis. *Psychological Methods*, 2, 192–199.
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent groups designs. *Psychological Methods*, 7, 105–125.
- Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika*, 8, 161–168.
- Mosteller, F., & Bush, R. R. (1954). Selected quantitative techniques. In Lindzey, G. (Ed.), *Handbook of social psychology: Vol. I. Theory and method*. Cambridge, MA: Addison-Wesley.
- Mosteller, F., & Colditz, G. A. (1996). Understanding research synthesis (meta-analysis). *Annual Review of Public Health*, 17, 1–17.
- Mosteller, F., & Moynihan, D. (1972). *On equality of educational opportunity*. New York: Vintage.
- Mount, M. K., & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research

and practice in human resources management. In Ferris, G. R. (Ed.), *Research in personnel and human resources management* (Vol. 13, pp. 153–200). Greenwich, CT: JAI Press.

Mullen, B. (1989). *Advanced BASIC meta-analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Mullen, B., & Rosenthal, R. (1985). *BASIC meta-analysis: Procedures and programs*. Hillsdale, NJ: Lawrence Erlbaum.

Murphy, K. R. (1997). Meta-analysis and validity generalization. In Anderson, N., & Herriott, P. (Eds.), *International handbook of selection and assessment* (pp. 323–342). Chichester, UK: Wiley.

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

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

Myers, D. G. (1991). Union is strength: A consumer's view of meta-analysis. *Personality and Social Psychology Bulletin*, 17, 265–266.

Nancy-Universite. (2008, October). *International workshop on meta-analysis in economics and business*, Nancy, France.

Nathan, B. R., & Alexander, R. A. (1988). A comparison of criteria for test validation: A meta-analytic investigation. *Personnel Psychology*, 41, 517–535.

National Research Council. (1992). *Combining information: Statistical issues and opportunities for research*. Washington, DC: National Academy of Sciences Press.

Nicol, T. S., & Hunter, J. E. (1973, August). *Mathematical models of the reliability of the semantic differential*. Paper presented at the Psychometric Society, Chicago.

Nouri, H., & Greenberg, R. H. (1995). Meta-analytic procedures for estimation of effect sizes in experiments using complex analysis of variance. *Journal of Management*, 21, 801–812.

Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.

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

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science*, 7, 384–403.



Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: John Wiley.

Oh, I. S. (2007). In search of ideal methods of research synthesis over 30 years (1977-2006): Comparison of Hunter-Schmidt meta-analysis methods with other methods and recent improvements. *International Journal of Testing*, 7, 89–93.

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

Oh, I. S., Postlethwaite, B. E., & Schmidt, F. L. (2013). Rethinking the validity of interviews for employment decision making: Implications of recent developments in meta-analysis. In Svyantek, D. J., & Mahoney, K. (Eds.), *Received wisdom, kernels of truth, and boundary conditions in organizational studies* (pp. 297–329). New York: Information Age Publishing.

Olson, C. M., Rennie, D., Cook, D., Dickersin, K., Flanagan, A., Hogan, J. W., et al. (2002). Publication bias in editorial decision making. *Journal of the American Medical Association*, 287, 2825–2828.

Ones, D. S., & Viswesvaran, C. (2003). Job-specific applicant pools and national norms for personality scales: Implications for range restriction corrections in validation research. *Journal of Applied Psychology*, 88, 570–577.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology Monograph*, 78, 679–703.

Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2012). Integrity tests predict counterproductive work behaviors and job performance well: Comment on Van Iddekinge, Roth, Raymark, and Odle-Dusseau (2012). *Journal of Applied Psychology*, 97, 537–542.

Orlitzky, M. (2011). How can significance tests be deinstitutionalized? *Organizational Research Methods*, 20, 1–30.

Orwin, R. G. (1983). A fail-safe *N* for effect size. *Journal of Educational Statistics*, 8, 147–159.

Orwin, R. G., & Cordray, D. S. (1985). Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. *Psychological Bulletin*, 97, 134–147.

Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 177–206). New York: Russell Sage.



- Osburn, H. G. (1978). Optimal sampling strategies for validation studies. *Journal of Applied Psychology*, 63, 602–608.
- Osburn, H. G., & Callender, J. C. (1990). Accuracy of the validity generalization sampling variance estimate: A reply to Hoben Thomas. *Journal of Applied Psychology*, 75, 328–333.
- 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.
- Osburn, H. G., Callender, J. C., Greener, J. M., & Ashworth, S. (1983). Statistical power of tests of the situational specificity hypothesis in validity generalization studies: A cautionary note. *Journal of Applied Psychology*, 68, 115–122.
- Overton, R. C. (1998). A comparison of fixed effects and mixed (random effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods*, 3, 354–379.
- Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin*, 97, 307–315.
- Palmer, T. M., Peters, J. L., Sutton, A. J., & Moreno, S. G. (2008). Contour-enhanced funnel plots for meta-analysis. *The Stata Journal*, 8, 242–254.
- Payne, S. C., Youngcourt, S. S., & Beaubien, J. M. (2007). A meta-analytic examination of the goal orientation nomological net. *Journal of Applied Psychology*, 92(1), 128–150.
- Pearlman, K. (1982). *The Bayesian approach to validity generalization: A systematic examination of the robustness of procedures and conclusions*. Unpublished doctoral dissertation, Department of Psychology, George Washington University, Washington, DC.
- Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. *Journal of Applied Psychology*, 65, 373–406.
- Pearson, E. S. (1938). The probability integral transformation for testing goodness of fit and combining tests of significance. *Biometrika*, 30, 134–148.
- Peters, J. L., Sutton, A. J., Jones, D. R., & Abrams, K. R. (2010). Assessing publication bias in meta-analysis in the presence of between-study heterogeneity. *Journal of the Royal Statistical Society*, 173(Pt. 3), 575–591.
- Peters, L. H., Harthe, D., & Pohlman, J. (1985). Fiedler's contingency theory of leadership: An application of the meta-analysis procedures of Schmidt and Hunter. *Psychological Bulletin*, 97, 274–285.
- Peterson, N. G. (1982, October). *Investigation of validity generalization in clerical and technical/professional*

*occupations in the insurance industry*. Paper presented at the Conference on Validity Generalization, Personnel Testing Council of Southern California, Newport Beach.

Peto, R. (1987). Why do we need systematic overviews of randomized trials? *Statistics in Medicine*, 6, 233–240.

Pinello, D. R. (1999). Linking party to judicial ideology in American courts: A meta-analysis. *The Justice System Journal*, 20, 219–254.

Powell, K. S., & Yalcin, S. (2010). Managerial training effectiveness: a meta-analysis 1952-2002. *Personnel Review*, 39(2), 227–241.

Premack, S., & Wanous, J. P. (1985). Meta-analysis of realistic job preview experiments. *Journal of Applied Psychology*, 70, 706–719.

Pritchard, R. D., Harrell, M. M., DiazGranadaos, D., & Guzman, M. J. (2008). The productivity measurement and enhancement system: a meta-analysis. *Journal of Applied Psychology*, 93(3), 540–567.

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

Raju, N. S., & Brand, P. A. (2003). Determining the significance of correlations corrected for unreliability and range restriction. *Applied Psychological Measurement*, 27, 52–72.

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

Raju, N. S., Burke, M. J., & Normand, J. (1983). *The asymptotic sampling distribution of correlations corrected for attenuation and range restriction*. Unpublished manuscript, Department of Psychology, Illinois Institute of Technology, Chicago.

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

Raju, N. S., & Drasgow, F. (2003). Maximum likelihood estimation in validity generalization. In Murphy, K. R. (Ed.), *Validity generalization: A critical review*. Hillsdale, NJ: Lawrence Erlbaum.

Raju, N. S., & Fleer, P. G. (1997). *MAIN: A computer program for meta-analysis*. Chicago: Illinois Institute of Technology.

Raju, N. S., Fralicx, R., & Steinhaus, S. D. (1986). Covariance and regression slope models for studying validity generalization. *Applied Psychological Measurement*, 10, 195–211.

Raudenbush, S. W. (1994). Random effects models. In Cooper, H., & Hedges, L. V. (Eds.), *The handbook of research synthesis* (pp. 301–322). New York: Russell Sage.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random effects models. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 295–315). New York: Russell Sage.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Application and data analysis methods*. Thousand Oaks, CA: Sage.

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>

Reed, J. G., & Baxter, P. M. (2009). Using reference databases. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 73–102). New York: Russell Sage.

Renkewitz, F., Fuchs, H. M., & Fiedler, S. (2011). Is there evidence of publication bias in JDM research? *Judgment and Decision Making*, 6, 870–881.

Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7, 331–363.

Riketta, M. (2008). The causal relation between job attitudes and job performance: A meta-analysis of panel studies. *Journal of Applied Psychology*, 93(2), 472–481.

Rodgers, R. C., & Hunter, J. E. (1986). *The impact of management by objectives on organizational productivity*. Unpublished manuscript, Management Department, University of Texas at Austin.

Rosenberg, M. S., Adams, D. C., & Gurevitch, J. (1997). *MetaWin: Statistical software for meta-analysis with resampling tests*. Sunderland, MA: Sinauer Associates.

Rosenthal, R. (1961, September). *On the social psychology of the psychological experiment: With particular reference to experimenter bias*. In H. W. Riecken (Chair), *On the social psychology of the psychological experiment*. Symposium conducted at the meeting of the American Psychological Association, New York.

Rosenthal, R. (1963). On the social psychology of the psychological experiment: The experimenter's hypothesis as unintended determinant of experimental results. *American Scientist*, 51, 268–283.

- Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin*, 85, 185–193.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Rosenthal, R. (1983). Assessing the statistical and social importance of the effects of psychotherapy. *Journal of Consulting and Clinical Psychology*, 51, 4–13.
- Rosenthal, R. (1984). *Meta-analysis procedures for social research*. Beverly Hills, CA: Sage.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (2nd ed.). Newbury Park, CA: Sage.
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences*, 3, 377–386.
- Rosenthal, R., & Rubin, D. B. (1979a). Comparing significance levels of independent studies. *Psychological Bulletin*, 86, 1165–1168.
- Rosenthal, R., & Rubin, D. B. (1979b). A note on percent variance explained as a measure of the importance of effects. *Journal of Applied Psychology*, 64, 395–396.
- Rosenthal, R., & Rubin, D. B. (1982a). Comparing effect sizes of independent studies. *Psychological Bulletin*, 92, 500–504.
- Rosenthal, R., & Rubin, D. B. (1982b). Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology*, 74, 708–712.
- Rosenthal, R., & Rubin, D. B. (1982c). A simple, general purpose display of magnitude of experiment effect. *Journal of Educational Psychology*, 74, 166–169.
- Rosenthal, R., & Rubin, D. B. (1983). Ensemble-adjusted *p* values. *Psychological Bulletin*, 94, 540–541.
- Roth, P. L. (2008). Software review: Hunter-Schmidt Meta-Analysis Programs 1.1. *Organizational Research Methods*, 11, 192–196.
- Roth, P. L., BeVier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54, 297–330.
- Roth, P. L., BeVier, C. A., Switzer, F. S., & Shippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology*, 81, 548–556.

Roth, P. L., Bobko, P., Switzer, F. S., & Dean, M. A. (2001). Prior selection causes biased estimates of standardized ethnic group differences: Simulation and analysis. *Personnel Psychology, 54*, 297–330.

Rothstein, H. R. (1990). Interrater reliability of job performance ratings: Growth to asymptote level with increasing opportunity to observe. *Journal of Applied Psychology, 75*, 322–327.

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 Murphy, K. R. (Ed.), *Validity generalization: A critical review* (pp. 115–154). Mahwah, NJ: Lawrence Erlbaum.

Rothstein, H. R. (2008). Publication bias is a threat to the validity of meta-analytic results. *Journal of Experimental Criminology, 4*, 61–81.

Rothstein, H. R. (2012). Accessing relevant literature. In Cooper, H. (Ed.), *APA Research Methods in Psychology: Vol. 1. Foundations, planning, measures, and psychometrics*. Washington, DC: American Psychological Association.

Rothstein, H. R., & Hopewell, S. (2009). Grey literature. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 103–126). New York: Russell Sage.

Rothstein, H. R., McDaniel, M. A., & Borenstein, M. (2001). Meta-analysis: A review of quantitative cumulation methods. In Schmitt, N., & Drasgow, F. (Eds.), *Advances in measurement and data analysis*. San Francisco: Jossey-Bass.

Rothstein, H. R., Schmidt, F. L., Erwin, F. W., Owens, W. A., & Sparks, C. P. (1990). Biographical data in employment selection: Can validities be made generalizable? *Journal of Applied Psychology, 75*, 175–184.

Rothstein, H. R., Sutton, A., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. London: Wiley.

Rubin, D. B. (1990). A new perspective on meta-analysis. In Wachter, K. W., & Straf, M. L. (Eds.), *The future of meta-analysis*. New York: Russell Sage.

Rubin, D. B. (1992). Meta-analysis: Literature synthesis or effect size surface estimation? *Journal of Educational Statistics, 17*, 363–374.

Sackett, P. R., Harris, M. M., & Orr, J. M. (1986). On seeking moderator variables in the meta-analysis of correlational data: A Monte Carlo investigation of statistical power and resistance to Type I error. *Journal of Applied Psychology, 71*, 302–310.

Sackett, P. R., Laczo, R. M., & Arvey, R. D. (2002). The effects of range restriction on estimates of criterion interrater reliability: Implications for validation research. *Personnel Psychology*, 55, 807–825.

Sanchez-Meca, J., Lopez-Lopez, J. A., & Lopez-Pina, J. A. (in press). Some recommended statistical analytic practices when reliability generalization studies are conducted. *British Journal of Mathematical and Statistical Psychology*.

Sanchez-Meca, J., & Marin-Martinez, F. (1998). Weighting by inverse variance or by sample size in meta-analysis: A simulation study. *Educational and Psychological Measurement*, 58, 211–220.

SAS Institute, Inc. (2003). *SAS/STAT software, Version 9.1*. Cary, NC: Author. <http://www.sas.com/>

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple study articles. *Psychological Methods*, 17, 551–566.

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, 31, 699–714.

Schmidt, F. L. (1972). The reliability of differences between linear regression weights in applied differential psychology. *Educational and Psychological Measurement*, 32, 879–886.

Schmidt, F. L. (1988). Validity generalization and the future of criterion-related validity. In Wainer, H., & Braun, H. (Eds.), *Test validity* (pp. 173–189). Hillsdale, 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. (2002). The role of general cognitive ability in job performance: Why there cannot be a debate. *Human Performance*, 15, 187–210.

Schmidt, F. L. (2003). John E. Hunter, 1939-2002. *American Psychologist*, 58, 238.

Schmidt, F. L. (2008). Meta-analysis: A constantly evolving research tool. *Organizational Research Methods*, 11, 96–113.

Schmidt, F. L. (2010). Detecting and correcting the lies that data tell. *Perspectives on Psychological Science*, 5, 233–242.

Schmidt, F. L. (in press). History of the development of the Schmidt-Hunter meta-analysis methods. *Research Synthesis Methods*.

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

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

Schmidt, F. L., & Hunter, J. E. (1978). Moderator research and the law of small numbers. *Personnel Psychology*, 31, 215–232.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128–1137.

Schmidt, F. L., & Hunter, J. E. (1984). A within setting test of the situational specificity hypothesis in personnel selection. *Personnel Psychology*, 37, 317–326.

Schmidt, F. L., & Hunter, J. E. (1992). Development of causal models of job performance. *Current Directions in Psychological Science*, 1, 89–92.

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1, 199–223.

Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In Harlow, L., Muliak, S., & Steiger, J. (Eds.), *What if there were no significance tests?* (pp. 37–64). Mahwah, NJ: Lawrence Erlbaum.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.

Schmidt, F. L., & Hunter, J. E. (1999a). Comparison of three meta-analysis methods revisited: An analysis of Johnson, Mullen, and Salas (1995). *Journal of Applied Psychology*, 84, 114–148.

Schmidt, F. L., & Hunter, J. E. (1999b). Theory testing and measurement error. *Intelligence*, 27, 183–198.

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

Schmidt, F. L., Hunter, J. E., & Caplan, J. R. (1981a). *Selection procedure validity generalization*



(transportability) results for three job groups in the petroleum industry. Washington, DC: American Petroleum Institute.

Schmidt, F. L., Hunter, J. E., & Caplan, J. R. (1981b). Validity generalization results for two job groups in the petroleum industry. *Journal of Applied Psychology*, 66, 261–273.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). The impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology*, 64, 609–626.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, 71, 432–439.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology*, 73, 46–57.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. M., & Trattner, M. H. (1986). The economic impact of job selection methods on the size, productivity, and payroll costs of the federal workforce: An empirical demonstration. *Personnel Psychology*, 39, 1–29.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences and validity of aptitude tests in selection: A red herring. *Journal of Applied Psychology*, 66, 166–185.

Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1982). Progress in validity generalization: Comments on Callender and Osburn and further developments. *Journal of Applied Psychology*, 67, 835–845.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Caplan, J. R. (1981). *Validity generalization results for three occupations in Sears, Roebuck and Company*. Chicago: Sears, Roebuck and Company.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Hirsh, H. R. (1985). Forty questions about validity generalization and meta-analysis. *Personnel Psychology*, 38, 697–798.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. (1979). Further tests of the Schmidt-Hunter Bayesian validity generalization procedure. *Personnel Psychology*, 32, 257–381.

Schmidt, F. L., Hunter, J. E., & Raju, N. S. (1988). Validity generalization and situational specificity: A second look at the 75% rule and the Fisher's  $z$  transformation. *Journal of Applied Psychology*, 73, 665–672.

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



Schmidt, F. L., & Kaplan, L. B. (1971). Composite vs. multiple criteria: A review and resolution of the controversy. *Personnel Psychology*, 24, 419–434.

Schmidt, F. L., Law, K. S., Hunter, J. E., Rothstein, H. R., Pearlman, K., & McDaniel, M. (1989). *Refinements in validity generalization methods (including outlier analysis)*. Unpublished paper, Department of Management and Organization, University of Iowa, Iowa City.

Schmidt, F. L., Law, K. S., Hunter, J. E., Rothstein, H. R., Pearlman, K., & McDaniel, M. (1993). Refinements in validity generalization methods: Implications for the situational specificity hypothesis. *Journal of Applied Psychology*, 78, 3–13.

Schmidt, F. L., & Le, H. (2004). *Software for the Hunter-Schmidt meta-analysis methods (versions 1.0 and 1.1)*. Iowa City: University of Iowa, Department of Management & Organizations.

Schmidt, F. L., & Le, H. (2014). *Software for the Hunter-Schmidt meta-analysis methods (version 1.2)*. Iowa City: University of Iowa Department of Management & Organizations.

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond Alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods*, 8, 206–234.

Schmidt, F. L., Le, H., & Oh, I.-S. (2009). Correcting for the distorting effects of study artifacts in meta-analysis. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 317–334). New York: Russell Sage.

Schmidt, F. L., Le, H., & Oh, I.-S. (in press). Are true scores and construct scores the same? A critical examination of their substitutability and the implications for research results. *International Journal of Selection and Assessment*.

Schmidt, F. L., Ocasio, B. P., Hillery, J. M., & Hunter, J. E. (1985). Further within-setting empirical tests of the situational specificity hypothesis in personnel selection. *Personnel Psychology*, 38, 509–524.

Schmidt, F. L., & Oh, I.-S. (2013). Methods for second order meta-analysis and illustrative applications. *Organizational Behavior and Human Decision Making*, 121, 204–218.

Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed vs. random models in meta-analysis: Model properties and comparison of differences in results. *British Journal of Mathematical and Statistical Psychology*, 62, 97–128.

Schmidt, F. L., Oh, I.-S., & Le, H. (2006). Increasing the accuracy of corrections for range restriction:

Implications for selection procedure validities and other research results. *Personnel Psychology*, 59, 281–305.

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. *Personnel Psychology*, 33, 705–724.

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

Schmidt, F. L., & Rothstein, H. R. (1994). Application of validity generalization methods of meta-analysis to biographical data scores in employment selection. In Stokes, G. S., Mumford, M. D., & Owens, W. A. (Eds.), *The biodata handbook: Theory, research, and applications* (pp. 237–260). Chicago: Consulting Psychologists Press.

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

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

Schmidt, F. L., & Zimmerman, R. (2004). A counter-intuitive hypothesis about interview validity and some supporting evidence. *Journal of Applied Psychology*, 89, 553–561.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta-analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*, 37, 407–422.

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge, MA: Hogrefe and Huber.

Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Journal of Psychology*, 215, 90–103.

Schwab, D. P., Olian-Gottlieb, J. D., & Heneman, H. G., III (1979). Between subject's expectancy theory research: A statistical review of studies predicting effort and performance. *Psychological Bulletin*, 86, 139–147.

Schwarzer, G. (2010). *Meta: Meta-analysis with R (R package version 1.6-0)*. <http://CRAN.R-project.org/package=meta>

Schwarzer, G., Carpenter, J., & Rucker, G. (2010). Empirical evaluation suggests Copas selection model is preferable to trim-and-fill method for selection bias in meta-analysis. *Journal of Clinical Epidemiology*, 63,

282–288.

Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, 309–316.

Shadish, W. R. (1996). Meta-analysis and the exploration of causal mediating processes: A primer of examples, methods, and issues. *Psychological Methods*, 1, 47–65.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shadish, W. R., Matt, G. E., Navarro, A. M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin*, 126, 512–529.

Shadish, W. R., & Ragsdale, K. (1996). Random versus nonrandom assignment in psychotherapy experiments: Do you get the same answer? *Journal of Consulting and Clinical Psychology*, 64, 1290–1305.

Shadish, W. R., Robinson, L., & Lu, C. (1999). *ES: A computer program and manual for effect size calculation*. St. Paul, MN: Assessment Systems Corporation.

Shah, B. V., Barnwell, B. G., & Bieler, G. S. (1995). *SUDAAN user's manual: Software for analysis of correlated data*. Research Triangle Park, NC: Research Triangle Institute.

Sharf, J. (1987). Validity generalization: Round two. *The Industrial-Organizational Psychologist*, 25, 49–52.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (in press). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General*.

Sitzmann, T., Brown, K. G., Casper, W. J., Ely, K., & Zimmerman, R. D. (2008). A review and meta-analysis of the nomological network of trainee reactions. *Journal of Applied Psychology*, 93(2), 280–295.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *The Educational Researcher*, 15, 5–11.

Smith, M., & Glass, G. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal*, 17, 419–433.

Smith, M. L. (1980). Sex bias in counseling and psychotherapy. *Psychological Bulletin*, 87, 392–407.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752–760.

Smithson, M. (2000). *Statistics with confidence*. London: Sage.

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement*, 61, 605–632.

Snedecor, G. W. (1946). *Statistical methods* (4th ed.). Ames: Iowa State College Press.

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

Spector, P. E., & Levine, E. L. (1987). Meta-analysis for integrating study outcomes: A Monte Carlo study of its susceptibility to Type I and Type II errors. *Journal of Applied Psychology*, 72, 3–9.

SPSS, Inc. (2006). *SPSS for Windows, Release 15*. Chicago: Author. <http://www.spss.com>

Stanley, J. C. (1971). Reliability. In Thorndike, R. L. (Ed.), *Educational measurement* (2nd ed., pp. 356–442). Washington, DC: American Council on Education.

Stanley, T. D. (1998). New wine in old bottles: A meta-analysis of Ricardian equivalence. *Southern Economic Journal*, 64, 713–727.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of Economic Perspectives*, 15, 131–150.

Stanley, T. D., & Jarrell, S. B. (1989). Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys*, 3, 161–169.

Stanley, T. D., & Jarrell, S. D. (1998). Gender wage discrimination bias? A metaregression analysis. *Journal of Human Resources*, 33, 947–973.

Stanley, T. D., Jarrell, S. D., & Doucouliagos, H. (2010). Could it be better to discard 90% of the data? A statistical paradox. *American Statistician*, 64, 70–77.

StataCorp. (2007). *Stata statistical software: Release 9.2*. College Station, TX: Author. <http://www.stata.com>

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

Sterling, T. C. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance or vice versa. *Journal of the American Statistical Association*, 54, 30–34.

Sterling, T. C., Rosenbaum, W., & Weinkam, J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician*, 49, 108–112.

Sterne, J. A. C. (Ed.). (2009). *Meta-analysis in Stata: An updated collection from the Stata Journal*. College Station, TX: Stata Press.

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–110). New York: John Wiley.

Stevens, J. P. (1984). Outliner and influential data points in regression analysis. *Psychological Bulletin*, 95, 334–344.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams, R. M., Jr. (1949). *The American soldier: Adjustment during Army life* (Vol. 1). Princeton, NJ: Princeton University Press.

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, 7, 670–688.

Strube, M. J. (1988). Averaging correlation coefficients: Influence of heterogeneity and set size. *Journal of Applied Psychology*, 73, 559–568.

Sutton, A. J. (2005). Evidence concerning the consequences of publication and related biases. In Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 175–192). West Sussex, UK: Wiley.

Sutton, A. J. (2009). Publication bias. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (2nd ed., pp. 435–452). New York: Russell Sage.

Tamrakar, C. (2012). *Relationship of satisfaction and loyalty across industry and geographical regions: A meta-analysis*. Unpublished paper, Marketing Department, University of Iowa, Iowa City.

Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the impact of *Culture's Consequences*: A three-decade, multilevel, meta-analytic review of Hofstede's cultural value dimensions. *Journal of Applied Psychology*, 95(3), 405–439.

Taveggia, T. (1974). Resolving research controversy through empirical cumulation. *Sociological Methods and*

*Research*, 2, 395–407.

Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009). Transfer of management training from alternative perspectives. *Journal of Applied Psychology*, 94(1), 104–121.

Terborg, J. R., & Lee, T. W. (1982). Extension of the Schmidt-Hunter validity generalization procedure to the prediction of absenteeism behavior from knowledge of job satisfaction and organizational commitment. *Journal of Applied Psychology*, 67, 280–296.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2002, May 10). *Adjusting for publication bias in the presence of heterogeneity*. Paper presented at the Meta-Analysis Symposium, Mathematical Research Institute, University of California, Berkeley.

Thomas, H. (1988). What is the interpretation of the validity generalization estimate  $Sp^2 = Sr^2 - Sc^2$ ? *Journal of Applied Psychology*, 73, 679–682.

Thomas, K. (2013, June 29). Breaking the seal on drug research. *New York Times*.

Thompson, B. (2002, April). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, pp. 25–32.

Thompson, W. A. (1962). The problem of negative estimates of variance components. *Annals of Mathematical Statistics*, 33, 273–289.

Thorndike, R. L. (1933). The effect of the interval between test and retest on the constancy of the IQ. *Journal of Educational Psychology*, 25, 543–549.

Thorndike, R. L. (1949). *Personnel selection*. New York: John Wiley.

Thorndike, R. L. (1951). Reliability. In Lindquist, E. F. (Ed.), *Educational measurement* (pp. 560–620). Washington, DC: American Council on Education.

Tracz, S. M., Elmore, P. B., & Pohlmann, J. T. (1992). Correlational meta-analysis: Independent and nonindependent cases. *Educational and Psychological Measurement*, 52, 879–888.

Trafimow, D., MacDonald, J. A., Rice, S., & Clason, D. L. (2010). How often is *p-rep* close to the true replication probability? *Psychological Methods*, 15, 300–307.

Trikalinos, T. A., & Ioannidis, J. P. (2005). Assessing the evolution of effect sizes over time. In Rothstein, H. R., Sutton, J. A., & Borenstein, M. (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 241–260). New York: John Wiley.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In Olkin, I., Ghurye, J. G., Hoeffding, W., Madoo, W. G., & Mann, H. (Eds.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.

Turner, R. M., Spiegelhalter, D. J., Smithe, G. C. S., & Thompson, S. G. (2009). Bias modeling in evidence synthesis. *Journal of the Royal Statistical Society*, 172(Pt. 1), 21–47.

*United States v. City of Torrance*, 163F.R.D. 590 (C.D. Cal., 1995).

Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development*, 44, 159–168.

Valentine, J. C. (2009). Judging the quality of primary research. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 129–146). New York: Russell Sage.

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, 96, 1167–1194.

Vevea, J. L., & Citkowicz, M. (2008, July). *Inference and estimation using conditionally random models: A Monte Carlo*. Paper presented at the third annual conference of the Society for Research Synthesis Methods, Corfu, Greece.

Vevea, J. L., Clements, N. C., & Hedges, L. V. (1993). Assessing the effects of selection bias on validity data for the General Aptitude Test Battery. *Journal of Applied Psychology*, 78, 981–987.

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60, 419–435.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, 10, 428–443.

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, 36, 1–42.

Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equation modeling. *Personnel Psychology*, 48, 865–885.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, 81, 557–560.



Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2002). The moderating influence of job performance dimensions on convergence of supervisory and peer ratings of job performance. *Journal of Applied Psychology, 87*, 345–354.

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in job performance ratings? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*, 108–131.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4*, 274–290.

Wachter, K. W., & Straf, M. L. (Eds.). (1990). *The future of meta-analysis*. New York: Russell Sage.

Whetzel, D. L., & McDaniel, M. A. (1988). Reliability of validity generalization data bases. *Psychological Reports, 63*, 131–134.

White, H. D. (2009). Scientific communication and literature retrieval. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 51–74). New York: Russell Sage.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology, 75*, 315–321.

Whitman, D. S., Van Rooy, D. L., & Viswesvaran, C. (2010). Satisfaction, citizenship behaviors, and performance in work units: A meta-analysis of collective construct relations. *Personnel Psychology, 63*, 41–81.

Wilkinson, L., & The APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604. (Reprint available through the APA home page: <http://www.apa.org/journals/amp/amp548594.html>)

Wilson, D. B. (2009). Systematic coding. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 159–176). New York: Russell Sage.

Wolins, L. (1962). Responsibility for raw data. *American Psychologist, 17*, 657–658.

Wortman, P. M. (1983). Evaluation research: A methodological perspective. *Annotated Review of Psychology, 34*, 223–260.

Wortman, P. M. (1994). Judging research quality. In Cooper, H., & Hedges, L. V. (Eds.), *Handbook of research*



*synthesis* (pp. 97–110). New York: Russell Sage.

Wortman, P. M., & Bryant, F. B. (1985). School desegregation and black achievement: An integrative review. *Sociological Methods and Research*, 13, 289–324.

Yong, E. (2012, July 12). Uncertainty shrouds psychologist's resignation: Lawrence Sanna departed University of Michigan amid questions over his work from 'data detective' Uri Simonsohn. *Nature/News*, pp. 2–4.

Zimmer, C. (2012a, October 1). Misconduct widespread in retracted science papers, study finds. *New York Times*.

Zimmer, C. (2012b, April 16). A sharp rise in retractions prompts calls for reform. *New York Times*.

Zimmerman, R.D. (2008). Understanding the impact of personality traits on individuals' turnover decisions: a meta-analytic path model. *Personnel Psychology*, 61(2), 309–348.

# Author Index

- Abrams, K. R., [514](#)
- Adams, D. C., [479](#)
- Aguinis, H., [35](#), [98](#), [193–194](#), [238](#), [375](#), [389](#), [471](#), [475](#), [491](#), [530](#), [554](#)
- Aiken, L. S., [472](#)
- Albright, L. E., [376](#)
- Alexander, R. A., [48–50](#), [139](#), [195–197](#), [241](#), [439](#)
- Allen, M., [253](#), [360](#)
- Alliger, G. M., [27](#), [48–50](#), [195–196](#)
- Aloe, A. M., [386](#)
- Altman, D. G., [228](#), [520](#)
- American Psychological Association (APA), [14](#), [20](#), [490](#), [493](#)
- Anselmi, T. V., [167](#)
- Antman, E. M., [32](#)
- APA Task Force on Statistical Inference, [14](#)
- Aronson, E., [373](#)
- Arthur, W., [27](#), [231](#), [236](#), [365](#), [482](#)
- Arvey, R. D., [29](#), [191](#), [345](#), [556](#)
- Ashworth, S., [375](#)
- Asplund, J., [26](#)
- Auerbach, H. A., [68](#), [253](#)
- Aytug, Z. G., [406](#), [490](#), [493](#)
- Baker, R., [236](#)
- Bakker, M., [524–525](#)
- Bangert-Drowns, R. L., [37](#), [456](#), [467–469](#)
- Banks, G. C., [406](#), [529–530](#)
- Banks, K. P., [530](#)
- Barnett, V., [235](#)
- Barnwell, B. G., [448](#)
- Barrick, M. R., [188](#), [197](#)
- Bartlett, C. J., [458](#)
- Baxter, P. M., [484](#)
- Beal, D. J., [236](#)
- Beaubien, J. M., [28](#)

- Becker, B. J., [25](#), [234](#), [352–353](#), [386](#), [456](#), [533](#)
- Bedeian, A. G., [530](#)
- Begg, C. B., [457](#), [513](#), [533](#), [537](#), [541](#)
- Bell, S. T., [27](#)
- Benjamin, A. J., [423](#)
- Bennett, W., [27](#), [482](#)
- Bentson, C., [188](#)
- Berlin, J. A., [520](#), [550](#)
- Bernstein, I. H., [35](#), [399](#), [401](#), [404](#), [475](#), [554](#)
- Bettencourt, B. A., [423](#)
- BeVier, C. A., [188](#), [290](#)
- Bieler, G. S., [448](#)
- Bijmolt, T. H. A., [440](#), [448](#)
- Billings, R., [232](#)
- Bjornstad, J. F., [440](#)
- Bloom, B. S., [463](#)
- Bobko, C., [289](#)
- Bobko, P., [26](#), [132](#), [145](#), [289–290](#), [293](#), [372–373](#), [458](#)
- Bodner, T. E., [29](#)
- Bond, C. F., Jr., [29](#)
- Bonett, D. G., [95](#), [221](#), [425](#), [428](#)
- Bono, J. E., [28](#), [93](#), [230](#)
- Borenstein, M., [14](#), [33](#), [88](#), [121](#), [159](#), [225](#), [228](#), [280](#), [345](#), [347](#), [350](#), [363](#), [367](#), [371](#), [426](#), [428](#), [514](#), [540](#), [550](#)
- Borman, G. D., [490](#)
- Bosco, F. A., [491](#), [530](#)
- Botella, J., [168](#)
- Bozarth, J. D., [527](#)
- Brand, P. A., [131](#)
- Brannick, M. T., [35](#), [95](#), [147](#), [198](#), [221–224](#), [226](#), [365–366](#), [368](#), [421–422](#), [424–425](#), [471](#), [475](#), [554](#)
- Brennan, R. L., [393](#)
- Brogden, H. E., [182](#)
- Brown, K. G., [27](#)
- Brown, S. H., [163](#)
- Brozek, J., [458](#)
- Bryant, F. B., [469](#)

- Bryant, N. D., [47](#), [125](#), [157](#), [192](#)
- Bryk, A. S., [389](#)– [390](#)
- Burke, M. J., [27](#), [131](#), [158](#), [169](#), [172](#)– [173](#), [197](#)– [198](#), [223](#), [227](#), [230](#), [239](#), [344](#)– [345](#), [366](#), [474](#), [481](#)– [482](#)
- Bush, R. R., [455](#)
- Bushman, B. J., [455](#), [457](#)
- Busse, T. V., [469](#), [487](#)
- Cafri, G., [424](#)
- Callender, J. C., [45](#), [169](#), [172](#), [179](#), [183](#), [186](#), [197](#)– [198](#), [219](#), [223](#), [226](#)– [227](#), [230](#), [239](#), [366](#), [375](#), [392](#), [395](#), [421](#), [423](#), [464](#), [474](#)
- Campbell, D. T., [355](#)– [357](#), [360](#)– [361](#), [373](#), [485](#)
- Campbell, J. P., [529](#)
- Caplan, J. R., [195](#)
- Carey, B., [521](#), [524](#)
- Carlsmith, J., [373](#)
- Carlson, D. L., [10](#)
- Carlson, K. D., [29](#), [93](#), [157](#), [160](#), [188](#), [353](#)– [355](#)
- Carlson, M., [489](#)
- Carpenter, J., [546](#)
- Carr, L., [48](#)– [50](#)
- Carson, K. P., [48](#)– [50](#), [195](#)– [196](#)
- Carver, R. P., [9](#)– [10](#)
- Casper, W. J., [27](#)
- Cattin, P., [82](#), [385](#), [510](#)
- Chalmers, T., [31](#)– [32](#), [37](#)
- Chan, A. W., [520](#)
- Chan, D. K., [438](#)– [440](#)
- Chan, M. E., [29](#), [556](#)
- Chan, W., [125](#), [234](#), [472](#)– [473](#)
- Chase, L. J., [526](#)
- Chase, R. B., [526](#)
- Chelimsky, E., [33](#)
- Cheung, M. W. L., [234](#), [388](#), [391](#), [464](#), [472](#)– [473](#), [477](#)
- Cheung, S. F., [438](#)– [440](#)
- Chinn, S., [280](#), [428](#)
- Citkowicz, M., [391](#)

- Clarke, M., [490](#), [493](#)
- Clements, N. C., [516](#)
- Coffman, D. L., [232](#)
- Coggin, T. D., [27](#), [66](#)
- Cohen, J., [8–10](#), [18](#), [248](#), [284](#), [427](#), [472](#), [527](#)
- Cohen, P., [472](#)
- Colbert, A. E., [28](#)
- Colditz, G. A., [227](#), [236](#)
- Cole, D. A., [345](#)
- Coleman, J. S., [215](#)
- Collins, D. B., [27](#)
- Collins, J. M., [496](#)
- Colquitt, J. A., [27](#)
- Conda, X., [32](#)
- Cook, T. D., [24–25](#), [35](#), [42](#), [47](#), [107](#), [121](#), [222](#), [261](#), [276](#), [314](#), [355](#), [365](#), [373](#), [386](#), [475](#), [484](#), [488](#), [499](#), [554](#)
- Cooper, H., [29](#), [223](#), [367](#), [405](#), [454–455](#), [483–484](#), [486–489](#), [493](#), [519–520](#)
- Copas, J. B., [546](#)
- Cordray, D. S., [34](#), [110](#), [386](#), [539](#)
- Corey, D. M., [236](#)
- Cortina, J. M., [293](#), [344–345](#), [350](#), [352–353](#)
- Coursol, A., [513](#), [526](#)
- Coward, W. M., [458](#)
- Cronbach, L. J., [xxvii](#), [21](#), [79](#), [116](#), [118–119](#), [393](#)
- Cronshaw, S. F., [195–196](#)
- Cui, Y., [125](#)
- Cumming, G., [14](#)
- Cunha, P. V., [26](#)
- Cureton, E. E., [132](#)
- Dalton, C. M., [491](#), [530](#)
- Dalton, D. R., [491](#), [530](#)
- Dawes, R. M., [469](#)
- Day, R., [27](#)
- Dean, M. A., [26](#), [290](#)
- Deeks, J. J., [228](#)
- DeGeest, D. S., [xxvii](#)

- Demaree, R. G., [66](#)
- DeShon, R. P., [197](#), [345](#), [349–350](#), [352–353](#)
- DeVinney, L. C., [455](#)
- DiazGranados, D., [26](#)
- Dickersin, K., [513](#), [520](#), [550](#)
- Dieckmann, N. F., [29](#), [490](#)
- Donahue, W. A., [253](#), [360](#)
- Doucouliagos, C., [528–529](#)
- Drasgow, F., [169](#), [231](#), [388](#), [416](#), [464](#)
- Dunbar, S. B., [27](#), [124](#)
- Dunlap, W. P., [236](#), [288](#), [344–345](#), [479](#)
- Dunnette, M. D., [157](#), [378](#)
- Duval, S., [529](#), [538–539](#), [543–544](#)
- Dye, D., [157](#), [160](#), [188](#), [378](#)
- Eagly, A. H., [28](#), [454](#)
- Easterbrook, P. J., [520](#)
- Edens, P. S., [27](#)
- Egger, M., [535](#), [541](#)
- Ellsworth, P., [373](#)
- Elmore, P. B., [405](#)
- Ely, K., [27](#)
- Emerson, G. B., [526](#)
- Erlbaum, L., [479](#)
- Erlenmeyer-Kimling, L., [463](#)
- Erwin, F. W., [93](#)
- Fanelli, D., [523](#)
- Feldt, L. S., [132](#)
- Fiedler, K., [27](#), [524](#), [528](#)
- Fiedler, S., [530](#)
- Field, A. P., [221–223](#), [225–227](#), [365–366](#), [420](#), [422–423](#)
- Fisher, C. D., [26](#)
- Fisher, R. A., [10](#), [284](#), [463](#)
- Fleer, P. G., [158](#), [481](#)
- Fleishman, E. A., [463](#)
- Forsyth, R. A., [132](#)
- Fountoulakis, K. N., [32](#)

- Fralix, R., [219](#)
- Francis, G., [528](#), [542](#)
- Frantz, R. A., [32](#)
- Freund, P. A., [389](#)
- Fuchs, H. M., [530](#)
- Gaissmaier, W., [280](#)
- Gambara, H., [168](#)
- Gamez, W., [293](#)
- GAO, [33–34](#)
- Gardner, S., [32](#)
- Gast-Rosenberg, I., [169](#)
- Gaugler, B. B., [188](#), [381](#)
- Gendreau, P., [32](#)
- Gerbing, D. W., [24](#), [497](#)
- Gergen, K. J., [21](#)
- Gerhardt, M. W., [28](#)
- Geyskens, I., [26](#), [490–491](#), [493](#)
- Gherzi, D., [550](#)
- Ghiselli, E. E., [463](#), [516](#)
- Gigerenzer, G., [8–9](#), [18](#), [280](#), [428](#), [527](#)
- Gitelson, R., [26](#)
- Glass, G. V., [xxvi](#), [xxvii](#), [16–17](#), [21](#), [35](#), [94](#), [284](#), [286](#), [290](#), [349](#), [384](#), [387](#), [453–455](#), [463–469](#), [475](#), [477](#), [487–488](#), [493](#), [515](#), [518](#), [548](#), [555](#)
- Glennon, J. R., [376](#)
- Gleser, G. C., [393](#)
- Gleser, L. J., [448](#)
- Gokhale, S., [47](#), [125](#), [157](#), [192](#)
- Gonzales, M., [373](#)
- Gooding, R. Z., [383–384](#)
- Goodman, J. S., [167](#)
- Gopalan, R., [520](#)
- Gottfredson, L. S., [59](#)
- Gottfredson, R. K., [389](#)
- Gotzsche, P. C., [520](#)
- Green, B. F., [362](#)
- Greenberg, R. H., [345](#), [349](#), [353](#)

- Greener, J. M., [375](#)
- Greenhouse, J. B., [533](#), [535](#), [546](#)
- Greenwald, A. G., [526](#)
- Griffeth, R. W., [288](#)
- Grigg, J. A., [490](#)
- Grissom, R. J., [345](#)
- Gross, A. L., [126](#)
- Grubbs, F. E., [235](#)
- Gulliksen, H., [63](#), [235](#)– [236](#)
- Gurevitch, J., [479](#)
- Guttman, L., [9](#)
- Guzman, M., [26](#)
- Guzzo, R. A., [23](#)
- Haahr, M. T., [520](#)
- Hackman, J. R., [105](#)– [107](#), [216](#), [219](#)
- Haddock, C., [249](#), [280](#), [427](#)
- Hafdahl, A. R., [32](#), [221](#), [225](#)
- Hall, J. A., [362](#)
- Hall, S. M., [35](#), [95](#), [147](#), [198](#), [221](#)– [223](#), [226](#), [365](#)– [366](#), [422](#), [471](#), [475](#), [554](#)
- Halvorsen, K. T., [493](#)
- Hamilton, M. A., [250](#), [253](#)
- Hannan, R., [458](#)
- Harmon, C., [33](#)
- Harnisch, D. L., [27](#), [124](#)
- Harrell, M. M., [26](#)
- Harris, C., [525](#)
- Harris, M. M., [375](#)
- Hartanto, F., [345](#)
- Harter, J. K., [26](#), [116](#)
- Hartigan, J. A., [189](#)
- Hartke, D., [27](#)
- Hayes, T. L., [26](#), [29](#), [66](#), [95](#), [147](#), [222](#), [224](#)– [225](#), [227](#), [230](#), [241](#), [296](#), [365](#), [367](#)– [368](#), [398](#), [401](#), [416](#), [423](#)– [424](#), [470](#)– [471](#), [520](#), [541](#), [554](#), [558](#)
- Hazucha, J., [345](#)
- Hedges, L. V., [8](#), [16](#), [21](#), [42](#), [74](#), [88](#), [100](#), [112](#), [121](#)– [122](#), [144](#), [146](#)– [147](#), [159](#), [221](#)– [225](#), [230](#), [236](#), [241](#), [274](#), [276](#), [292](#)– [294](#), [345](#), [349](#), [351](#), [362](#), [365](#)– [367](#), [371](#), [375](#), [384](#), [387](#), [390](#)– [391](#), [414](#), [420](#)–



[426](#), [440](#), [448](#), [458–462](#), [464](#), [469–471](#), [477](#), [479](#), [516](#), [537](#), [539](#), [543–548](#), [551](#), [554–555](#)

- Heinsman, D. T., [279](#)
- Heneman, H. G., [III](#), [107](#)
- Higgins, J. P. T., [88](#), [228–229](#), [363](#)
- Hill, T. E., [76](#)
- Hillery, J. M., [12](#)
- Hirsh, H. R., [63](#), [79](#), [195](#), [276](#), [383](#)
- Hoffert, S. P., [33–34](#)
- Hofstede, G., [26](#)
- Holton, E. F., [27](#)
- Hopewell, S., [484](#)
- Hotelling, H., [66](#)
- Hoyt, W. T., [120](#)
- Hrobjartsson, A., [520](#)
- Huber, P. J., [235–236](#)
- Huffcutt, A. I., [231](#), [236](#), [365](#), [482](#)
- Hunt, M., [31](#), [33](#)
- Hunter, J. E., [xxiv](#), [xxvi](#), [xxviii](#), [xxxi–xxxii](#), [8–12](#), [18](#), [21](#), [23–27](#), [33](#), [40](#), [43–44](#), [46–47](#), [63–64](#), [66–69](#), [79–80](#), [95](#), [100–101](#), [116](#), [121](#), [124–126](#), [133](#), [135–136](#), [139–141](#), [144](#), [147](#), [154–157](#), [167–173](#), [188–189](#), [191–193](#), [195–199](#), [201](#), [215](#), [217](#), [221–225](#), [228–229](#), [233](#), [237](#), [239](#), [241–242](#), [250](#), [253](#), [257](#), [268](#), [272–273](#), [276](#), [282](#), [290](#), [296](#), [349](#), [360](#), [363](#), [365–368](#), [379](#), [381](#), [383](#), [386](#), [416](#), [420–421](#), [423–424](#), [427](#), [439](#), [441](#), [455–456](#), [463–464](#), [470–471](#), [476](#), [481–482](#), [487](#), [497](#), [499](#), [516](#), [519](#), [532](#)
- Hunter, R., [135](#), [188](#), [197](#), [199](#), [383](#)
- Ilies, R., [28](#), [116](#)
- Ioannidis, J., [525–526](#), [529](#), [542](#)
- Iyengar, S., [533](#), [535](#), [546](#)
- Jackson, C. L., [26](#)
- Jackson, D., [236](#)
- Jackson, G. B., [xxvi](#)
- Jackson, S. E., [23](#)
- James, L. R., [66](#)
- Jarrell, S. B., [33](#), [529](#)
- Jarvik, L. F., [463](#)
- Jensen, A. R., [214](#)
- Ji, F. X., [29](#)

- Johanssen-Schmidt, M. C., [28](#)
- John, L. K., [522](#), [524](#)– [527](#)
- Johnson, B. T., [344](#), [479](#)
- Johnson, M. C., [390](#)
- Jones, D. R., [514](#)
- Judge, T. A., [28](#), [93](#), [160](#), [230](#)
- Kahneman, D., [524](#)
- Kammeyer-Mueller, J. D., [387](#)– [388](#)
- Kaplan, L. B., [446](#)
- Karau, S. J., [28](#)
- Kasten, N., [389](#)
- Katzell, R. A., [23](#)
- Kelly, T. L., [132](#)
- Kemery, E. R., [288](#), [395](#)
- Kepes, S., [406](#), [528](#)– [530](#), [551](#)
- Kern, M. C., [406](#)
- Kilham, E. A., [26](#)
- Killeen, P. R., [10](#)
- Kim, J. J., [345](#)
- King, L. M., [63](#), [132](#), [276](#)
- Kirk, R. E., [14](#), [387](#)
- Kirkman, B. L., [26](#)
- Kirkpatrick, D. L., [27](#)
- Kirsch, M., [383](#)– [384](#)
- Kisamore, J. L., [224](#), [226](#), [368](#)
- Kline, R. B., [14](#)
- Koenka, A. C., [405](#)
- Kosalka, T., [28](#)
- Kotoz, R., [293](#)
- Krakovsky, M., [550](#)
- Kramer, G. G., [463](#)
- Krishnan, R., [26](#)
- Kulik, J. A., [469](#)
- Kulinskaya, E., [424](#)
- Kupelnick, B., [32](#)
- Kurz-Milcke, E., [280](#)

- Laczo, R. M., [191](#), [293](#)
- Landis, R. S., [234](#), [344](#)
- Landman, J. T., [469](#)
- Langlois, G. M., [158](#)
- Lanivich, S. E., [432](#)
- Lau, J., [32](#), [539](#)
- Lauver, K., [116](#)
- Law, K. S., [100](#), [102](#), [144](#), [146–147](#), [169–170](#), [184](#), [193–194](#), [222](#), [224](#), [227](#), [237](#), [240](#), [474](#)
- Lawshe, C. H., [376](#)
- Le, H., [xxiv](#), [16](#), [34](#), [40](#), [47](#), [92](#), [116–117](#), [119–120](#), [125](#), [146](#), [156–157](#), [166](#), [168–169](#), [173](#), [181](#), [185](#), [189](#), [192–193](#), [198](#), [276](#), [480](#), [486–487](#), [498](#), [553](#), [576](#)
- Lee, S., [489](#)
- Lee, T. W., [27](#)
- Lehrer, J., [526](#)
- Lent, R. H., [68](#), [253](#), [392](#), [517](#)
- Leone, F. C., [393](#)
- LePine, J. A., [26–27](#)
- Levin, L. S., [68](#), [253](#)
- Levine, E. L., [375](#), [394–395](#)
- Levine, E. N., [463](#)
- Levine, J. M., [463](#)
- Lewis, T., [235](#)
- Li, J. C., [125](#)
- Light, R. J., [454](#), [464](#), [493](#), [535](#)
- Linn, R. L., [27](#), [46](#), [124](#), [197–198](#)
- Lipsey, M. W., [246](#), [387](#), [480](#), [514](#), [518](#), [523](#)
- Lockhart, R. S., [14](#)
- Loewenstein, G., [522](#)
- Lopez-Lopez, J. A., [168](#)
- Lopez-Pina, J. A., [168](#)
- Lord, F., [35](#), [475](#), [554](#)
- Lu, C., [345](#)
- Lumley, T., [480](#)
- Mabe, P. A., [III](#), [27](#)
- MacCallum, R. C., [43](#), [133](#), [232](#), [268](#), [427](#)
- MacDonald, J. A., [10](#)

- MacMahon, S., [32](#), [110](#), [499](#)
- Magnusson, D., [399](#), [401](#), [404](#)
- Makhijani, M. G., [28](#)
- Malle, B. F., [29](#)
- Mann, C., [37](#)
- Mansfield, R. S., [469](#), [487](#)
- Marin-Martinez, F., [424](#)
- Martinussen, M., [440](#)
- Mathieu, J. E., [26](#)
- Matt, G. E., [35](#), [42](#), [47](#), [121](#), [223](#), [276](#), [314](#)
- Matthews, D. R., [520](#)
- Maurer, S. D., [188](#)
- Maxwell, S. E., [8](#), [519](#), [527](#)
- Mazumdar, M., [513](#), [541](#)
- McDaniel, M. A., [33](#), [188](#), [241](#)– [242](#), [380](#), [406](#), [445](#), [482](#), [489](#), [517](#), [528](#)– [529](#), [540](#)– [541](#)
- McGanney, M. L., [126](#)
- McGaw, B., [xxvii](#)
- McKenzie, R. C., [68](#), [215](#)
- McNatt, D. B., [534](#)
- McNemar, Q., [513](#)
- Meehl, P. E., [9](#), [21](#)
- Meinert, C., [520](#)
- Mendoza, J. L., [46](#)– [47](#), [124](#), [132](#), [154](#), [169](#), [172](#), [198](#)
- Miller, A. N., [530](#)
- Miller, N., [489](#)
- Millsap, R., [98](#), [193](#), [237](#)– [238](#)
- Min, Y., [520](#)
- Minder, C., [541](#)
- Moher, D., [31](#)
- Mondale, W., [19](#)
- Moreno, S. G., [537](#)
- Morgenthaler, S., [424](#)
- Morphy, P., [34](#), [386](#), [539](#)
- Morris, S. B., [345](#), [349](#)– [350](#), [352](#)– [353](#), [481](#)
- Mosier, C. I., [446](#)
- Mosier, S. B., [458](#)

- Mossholder, K. W., [395](#)
- Mosteller, F., [32](#), [215](#), [227](#), [236](#), [455](#)
- Mount, M. K., [188](#), [197](#)
- Moynihan, D., [215](#)
- Mulaik, S. A., [66](#)
- Muldrow, T. W., [68](#), [215](#)
- Mullen, B., [344](#)
- Mumford, M., [46–47](#), [124](#), [154](#), [172](#), [198](#)
- Murphy, K. R., [xxvii](#), [197](#), [372](#)
- Murphy, M. A., [188](#)
- Myers, D. G., [30](#)
- Nanda, H., [393](#)
- Nathan, B. R., [241](#), [439](#)
- National Research Council, [213](#), [224](#), [368](#), [375](#), [390–391](#), [414](#), [456](#), [458](#), [470–471](#)
- Navarro, A. M., [223](#)
- Nelson, L. D., [543](#)
- Nelson, L. S., [393](#)
- Nguyen, N. T., [541](#)
- Nicol, T. S., [257](#)
- Noe, R. A., [27](#), [383–384](#)
- Normand, J., [131](#), [158](#)
- Northrop, L. C., [195](#)
- Nouri, H., [345](#), [349–350](#), [352–353](#)
- Novick, M., [35](#), [475](#), [554](#)
- Nunnally, J., [35](#), [282](#), [399](#), [401](#), [404](#), [441](#), [475](#), [554](#)
- Nye, C. D., [388](#), [473](#)
- Oakes, M., [9–11](#)
- Ocasio, B. P., [12](#), [67](#), [379](#), [381](#), [522](#)
- Oh, I.-S., [16](#), [29](#), [34](#), [40](#), [66](#), [92](#), [95](#), [119](#), [146–147](#), [185](#), [189](#), [222](#), [224–225](#), [227](#), [230](#), [241](#), [276](#), [296](#), [365](#), [367–368](#), [396](#), [398](#), [401](#), [405–407](#), [416](#), [423–424](#), [470–471](#), [481](#), [499](#), [519–520](#), [529–530](#), [541](#), [549](#), [553–554](#), [558](#)
- Oldham, G. R., [216](#), [219](#)
- Olian-Gottlieb, J. D., [107](#)
- Olkin, I., [8](#), [31](#), [74](#), [146–147](#), [221](#), [223](#), [225](#), [274](#), [293–294](#), [366–367](#), [371](#), [384](#), [387](#), [391](#), [422](#), [448](#), [455](#), [458–462](#), [464](#), [469–471](#), [477](#), [479](#), [537](#), [539](#), [544–545](#), [555](#)
- Olson, C. M., [527](#)

- Ones, D. S., [24](#), [120](#), [188](#), [197](#), [229](#), [231–232](#), [386](#)
- Oosterbeck, H., [33](#)
- Orlitzky, M., [14](#)
- Orr, J. M., [375](#)
- Orwin, R. G., [110](#), [386](#), [488](#), [533](#)
- Osburn, H. G., [45–46](#), [169](#), [172](#), [179](#), [183](#), [186](#), [197–198](#), [223](#), [226–227](#), [230](#), [239](#), [366](#), [375](#), [392](#), [395](#), [421](#), [423](#), [464](#), [474](#)
- Outerbridge, A. N., [24–25](#), [64](#), [242](#)
- Overton, R. C., [222](#), [224](#), [241](#), [365](#), [368](#), [388](#), [390](#), [422–423](#)
- Owens, W. A., [93](#)
- Ozer, D. J., [214](#)
- Palmer, T. M., [537](#)
- Pashler, H., [525](#)
- Patton, G. K., [93](#)
- Payne, S. C., [28](#)
- Pearlman, K., [23](#), [79](#), [167](#), [169–170](#), [188](#), [194](#), [197](#), [379](#), [396](#), [454](#), [487](#), [516–517](#), [533](#)
- Pearson, E. S., [463](#)
- Peckham, P. D., [387](#)
- Peters, J. L., [514](#), [537–538](#)
- Peters, L. H., [27](#)
- Peterson, N. G., [157](#), [378](#)
- Peto, R., [32](#)
- Phillips, G., [223](#)
- Piccolo, R. F., [26](#), [28](#)
- Pierce, C. A., [35](#), [389](#), [491](#), [530](#)
- Pieters, R. G. M., [440](#), [448](#)
- Pigott, T. D., [88](#), [122](#), [159](#), [224](#), [367](#), [375](#), [386](#), [414](#), [470–471](#)
- Pillemer, D. B., [535](#)
- Pinello, D. R., [33](#)
- Pohlmann, J. T., [27](#), [405](#)
- Postlethwaite, B. E., [519](#)
- Postmes, T., [521–522](#)
- Powell, K. S., [27](#)
- Preacher, K. J., [43](#)
- Prelec, D., [522](#)
- Premack, S., [27](#)

- Pritchard, R. D., [26](#)
- Putka, D. J., [116](#), [432](#)
- Ragsdale, K., [279](#)
- Rajaratnam, N., [393](#)
- Raju, N. S., [23](#), [100](#), [131](#), [158](#), [167](#), [169](#), [172](#)– [173](#), [197](#)– [198](#), [219](#), [223](#), [225](#), [227](#), [229](#)– [231](#), [239](#), [366](#), [416](#), [420](#)– [421](#), [423](#)– [424](#), [464](#), [474](#), [481](#)
- Raudenbush, S. W., [229](#), [385](#), [389](#)– [390](#), [424](#)
- Reck, M., [188](#)
- Reed, J. G., [484](#)
- Reick, A., [132](#)
- Reinhardt, R. N., [169](#)
- Renkewitz, F., [530](#)
- Rice, S., [10](#)
- Richard, F. D., [29](#), [514](#)
- Riketta, M., [26](#)
- Rindskopf, D., [249](#)
- Roberts, R. R., [527](#)
- Robinson, L., [345](#)
- Rodgers, R. C., [116](#), [381](#), [383](#)
- Romashko, T., [463](#)
- Rosenbaum, W., [527](#)
- Rosenberg, M. S., [479](#)
- Rosenthal, D. B., [188](#)
- Rosenthal, R., [216](#), [223](#), [344](#), [362](#), [367](#)– [368](#), [371](#), [454](#)– [458](#), [463](#)– [464](#), [469](#)– [470](#), [515](#)– [516](#), [531](#), [533](#), [555](#)
- Roth, L., [395](#)
- Roth, P. L., [26](#), [188](#), [289](#)– [290](#), [432](#), [481](#)
- Rothstein, H. R., [33](#), [63](#), [88](#), [93](#), [120](#), [123](#), [157](#), [160](#), [188](#), [196](#), [228](#), [363](#), [396](#), [406](#), [478](#), [484](#), [514](#), [517](#), [521](#)
- Rounds, J., [388](#)
- Rubin, D. B., [xxvi](#), [36](#), [94](#), [159](#), [216](#), [223](#), [315](#), [344](#), [362](#), [367](#)– [368](#), [456](#)– [457](#), [464](#), [469](#)– [470](#), [475](#)– [476](#), [478](#), [555](#)– [556](#)
- Rucker, D. D., [43](#)
- Rucker, G., [546](#)
- Russ-Eft, D. F., [27](#)
- Sackett, P. R., [191](#), [293](#), [375](#), [458](#)

- Sanchez-Meca, J., [168](#), [424](#)
- Sanders, J. R., [387](#)
- SAS Institute, Inc., [480](#)
- Saul, J. R., [26](#)
- Schimmack, U., [528](#), [542](#)
- Schmid, C. H., [539](#)
- Schmidt, F. L., [xxiv](#), [xxvi](#), [xxvii](#), [xxviii](#), [xxxi](#), [8–12](#), [16](#), [18](#), [21](#), [23–26](#), [29](#), [32–34](#), [40](#), [43–44](#), [47](#), [63–64](#), [66–69](#), [78–80](#), [92–93](#), [95](#), [100–101](#), [116–117](#), [119–122](#), [125](#), [132–133](#), [136](#), [141](#), [144](#), [146–147](#), [155–156](#), [167–170](#), [172–173](#), [181–182](#), [185–186](#), [188–189](#), [192–199](#), [201](#), [215](#), [217](#), [220–225](#), [227–230](#), [235–242](#), [268](#), [276](#), [293](#), [296](#), [349](#), [353–355](#), [363](#), [365–368](#), [371–372](#), [377–379](#), [381](#), [384–386](#), [396](#), [398](#), [401](#), [405–407](#), [416](#), [420–421](#), [423–424](#), [427](#), [441](#), [446](#), [456](#), [463–464](#), [470–471](#), [476](#), [480–482](#), [487](#), [498](#), [516](#), [519–520](#), [522](#), [532](#), [541](#), [553–555](#), [558](#), [576](#)
- Schmitt, N., [383–384](#)
- Schneider, M., [541](#)
- Schram, C. M., [25](#), [456](#)
- Schulze, R., [95](#), [146–147](#), [221](#), [367](#), [421](#), [424–425](#)
- Schwab, D. P., [107](#)
- Schwartz, L. M., [280](#)
- Schwarzer, G., [480](#), [546](#)
- Scullen, S. E., [93](#)
- Sedlmeier, P., [8–9](#), [18](#), [527](#)
- Shadish, W. R., [25](#), [223](#), [249](#), [279](#), [345](#), [347–349](#), [352](#), [355](#), [367](#), [478–479](#)
- Shaffer, J. A., [189](#)
- Shah, B. V., [448](#)
- Shane, G. S., [79](#), [169](#), [188](#)
- Sharf, J., [189](#)
- Shi, J. Q., [546](#)
- Shippmann, J. S., [188](#)
- Shotland, A., [27](#)
- Simmons, J. P., [524–525](#), [543](#)
- Simonsohn, U., [543](#)
- Singer, J. D., [493](#)
- Sitzmann, T., [27](#)
- Slavin, R. E., [467](#), [484–485](#)
- Smith, G., [541](#)
- Smith, M. L., [xxvii](#), [284](#), [384](#), [387](#), [467](#), [514](#), [518](#)



- Smith, P., [32](#)
- Smith, P. V., [454](#), [464](#)
- Smith, W. J., [376](#)
- Smithe, G. C. S., [485](#)
- Smithson, M., [14](#)
- Snedecor, G. W., [469](#)
- Society for Industrial/Organizational Psychology (SIOP), [189](#)
- Sparks, C. P., [93](#)
- Spears, R., [521](#)– [522](#)
- Spector, P. E., [375](#), [394](#)– [395](#)
- Spiegelhalter, D. J., [229](#), [485](#)
- SPSS Inc., [480](#)
- Stafford, K. L., [132](#)
- Stanley, J. C., [116](#)– [117](#), [119](#), [355](#)– [357](#), [360](#)– [361](#), [485](#)
- Stanley, T. D., [33](#), [528](#)– [529](#), [537](#)
- Star, S. A., [455](#)
- StataCorp, [480](#)
- Staudte, R. G., [424](#)
- Stauffer, J. M., [132](#)
- Steel, P., [26](#)
- Steel, P. D., [387](#)– [388](#)
- Steenkamp, J. E. M., [26](#)
- Steinhaus, S. D., [219](#)
- Sterling, T. C., [527](#)
- Sterne, J. A. C., [480](#), [535](#), [541](#)
- Stock, W., [387](#)
- Stockman, D., [20](#)
- Stokes-Zoota, J. J., [29](#)
- Stone-Romero, E. F., [372](#)– [373](#)
- Stouffer, S. A., [455](#)
- Straf, M. L., [223](#), [367](#)
- Stroebe, W., [521](#)– [522](#)
- Strube, M. J., [220](#)– [221](#)
- Sturman, M. C., [35](#)
- Su, R., [388](#)
- Suchman, E. A., [455](#)

- Suero, M., [168](#)
- Sutton, A. J., [514](#), [537](#), [539](#)– [540](#), [551](#)
- Switzer, F. S., [III](#), [188](#), [290](#)
- Talley, A., [423](#)
- Tamrakar, C., [430](#)
- Tannenbaum, S. I., [27](#)
- Taras, V., [26](#)
- Taveggia, T., [22](#), [439](#)
- Taylor, H., [27](#)
- Taylor, P. J., [27](#)
- Taylor, S. G., [530](#)
- Terborg, J. R., [27](#)
- Terrin, N., [538](#)– [539](#)
- Thomas, A., [167](#)
- Thomas, H., [392](#)
- Thomas, K., [550](#)
- Thompson, B., [14](#), [168](#), [494](#)
- Thompson, S. G., [228](#)– [229](#), [363](#), [485](#)
- Thompson, W. A., [103](#), [393](#)
- Thorensen, C. J., [93](#)
- Thorndike, R. L., [116](#)– [117](#), [119](#), [123](#)– [124](#), [197](#), [463](#)
- Thornton, G. C., [188](#)
- Tiede, K., [458](#)
- Tipton, E., [390](#)
- Tracz, S. M., [405](#), [438](#)– [439](#)
- Trafimow, D., [10](#)
- Traver, H., [27](#)
- Trikalinos, T. A., [525](#), [542](#)
- Tukey, J. W., [63](#)– [64](#), [235](#)– [236](#)
- Turner, R. M., [485](#)
- Tweedie, R., [538](#)– [539](#), [543](#)
- Tyler, P., [290](#)
- Urry, V. E., [18](#)
- Vacha-Haase, T., [168](#), [494](#)
- Valentine, J., [423](#), [488](#)
- van Dijk, A., [524](#)– [525](#)

- van Engen, M. L., [28](#)
- Van Iddekinge, C. H., [432](#)
- Van Rooy, D. L., [26](#)
- Vaslow, J. B., [344](#)– [345](#)
- Vevea, J. L., [222](#)– [223](#), [225](#), [227](#), [230](#), [241](#), [365](#)– [368](#), [391](#), [420](#)– [425](#), [488](#), [516](#), [544](#), [547](#)– [549](#), [551](#)
- Viechtbauer, W., [480](#)
- Vieta, E., [32](#)
- Viswesvaran, C., [24](#), [26](#), [120](#), [123](#), [182](#), [188](#), [196](#)– [197](#), [229](#), [231](#)– [232](#), [364](#), [386](#), [455](#)
- Vul, E., [525](#)
- Wachter, K. W., [223](#), [367](#)
- Wagner, E. E., [513](#), [526](#)
- Walker, I., [33](#)
- Wang, M. D., [455](#), [457](#)
- Wanous, J. P., [27](#)
- Watson, D., [293](#)
- Weinkam, J., [527](#)
- West, S. G., [27](#), [472](#)
- Whetzel, D. L., [188](#), [406](#), [489](#), [517](#)
- White, H. D., [484](#)
- Whitehead, R., [98](#), [194](#), [238](#)
- Whitener, E. M., [228](#), [363](#)
- Whitman, D. S., [26](#)
- Wicherts, J. M., [524](#)– [525](#)
- Wigdor, A. K., [189](#)
- Wilkinson, L., [14](#)
- Willett, J. B., [493](#)
- Williams, M. A., [225](#)
- Williams, R. M., Jr., [455](#)
- Wilson, D. B., [246](#), [387](#), [480](#), [488](#), [514](#), [518](#), [523](#)
- Winkielman, P., [525](#)
- Wolins, L., [63](#), [235](#)
- Woloshin, S., [280](#)
- Woods, C. M., [544](#), [548](#)– [549](#), [551](#)
- Wortman, P. M., [467](#), [469](#), [488](#)
- Wroten, S., [232](#)
- Yalcin, S., [27](#)

- Yang, L.-Q., [424](#)
- Yong, E., [522](#), [524](#)
- Youngcourt, S. S., [28](#)
- Zhang, S., [43](#)
- Zhou, W., [406](#)
- Zimmer, C., [523](#)
- Zimmerman, R., [26–27](#), [241](#)