Copyright American Psychological Association

Contents

Volume 1: Foundations, Planning, Measures, and Psychometrics

Editorial Board xi
About the Editor-in-Chief xiii
About the Associate Editors xv
Contributors xix
A Note From the Publisher xxiii
Introduction: Objectives of Psychological Research and Their Relations to Research Methods xxv

Part I. Philosophical, Ethical, and Societal Underpinnings of Psychological Research. 1

Section 1. Philosophical Issues for Research in Psychology 3

Chapter 1. Perspectives on the Epistemological Bases for Qualitative Research 5
Carla Willig
Chapter 2. Frameworks for Causal Inference in Psychological Science 23
Peter M. Steiner, William R. Shadish, and Kristynn J. Sullivan

Section 2. Ethical and Professional Considerations in Conducting Psychological Research 57

Chapter 3. Ethics in Psychological Research: Guidelines and Regulations 59
Adam L. Fried and Kate L. Jansen
Chapter 4. Ethics and Regulation of Research With Nonhuman Animals 83
Sangeeta Panicker, Chana K. Akins, and Beth Ann Rice

Section 3. Cultural and Societal Issues in Conducting Psychological Research 95

Chapter 5. Cross-Cultural Research Methods. 97
David Matsumoto and Fons J. R. van de Vijver
Chapter 6. Research With Populations That Experience Marginalization 115
George P. Knight, Rebecca M. B. White, Stefanie Martinez-Fuentes, Mark W. Roosa, and Adriana J. Umaña-Taylor

Copyright American Psychological Association

Part II. Planning Research **133**

Chapter 7. Developing Testable and Important Research Questions 135
Frederick T. L. Leong, Neal Schmitt, and Brent J. Lyons
Chapter 8. Searching With a Purpose: How to Use Literature Searching to Support Your Research 151
Diana Ramirez and Margaret J. Foster
Chapter 9. Psychological Measurement: Scaling and Analysis 169
Heather Hayes and Susan E. Embretson
Chapter 10. Sample-Size Planning 189
Ken Kelley, Samantha F. Anderson, and Scott E. Maxwell
Chapter 11. Workflow and Reproducibility 211
Oliver Kirchkamp
Chapter 12. Obtaining and Evaluating Research Funding 227
Jonathan S. Comer and Amanda L. Sanchez

Part III. Measurement Methods **247**

Section 1. Behavior Observation **249**

Chapter 13. Behavioral Observation 251
Roger Bakeman and Vicenç Quera

Section 2. Self-Report **275**

Chapter 14. Question Order Effects 277
Lisa Lee, Parvati Krishnamurty, and Struther Van Horn
Chapter 15. Interviews and Interviewing Techniques 297
Anna Madill
Chapter 16. Using Intensive Longitudinal Methods in Psychological Research 327
Masumi Iida, Patrick E. Shrout, Jean-Philippe Laurenceau, and Niall Bolger
Chapter 17. Automated Analyses of Natural Language in Psychological Research 361
Laura K. Allen, Arthur C. Graesser, and Danielle S. McNamara

Section 3. Psychological Tests **381**

Chapter 18. Objective Tests as Instruments of Psychological Theory and Research . . . 383
David Watson
Chapter 19. Norm- and Criterion-Referenced Testing 407
Kurt F. Geisinger
Chapter 20. The Current Status of "Projective" "Tests" 433
Robert E. McGrath, Alec Twibell, and Elizabeth J. Carroll
Chapter 21. Brief Instruments and Short Forms 451
Emily A. Atkinson, Carolyn M. Pearson Carter, Jessica L. Combs Rohr, and Gregory T. Smith

Copyright American Psychological Association

Section 4. Chronometric and Psychophysical Measures 467

Chapter 22. Eye Movements, Pupillometry, and Cognitive Processes 469
Simon P. Liversedge, Sara V. Milledge, and Hazel I. Blythe
Chapter 23. Response Times 493
Roger Ratcliff
Chapter 24. Psychophysics: Concepts, Methods, and Frontiers 511
Allie C. Hexley, Takuma Morimoto, and Manuel Spitschan

Section 5. Measures in Psychophysiology 529

Chapter 25. The Perimetric Physiological Measurement of Psychological Constructs 531
Louis G. Tassinary, Ursula Hess, Luis M. Carcoba, and Joseph M. Orr
Chapter 26. Salivary Hormone Assays 565
Linda Becker, Nicolas Rohleder, and Oliver C. Schultheiss

Section 6. Measures in Neuroscience 579

Chapter 27. Electro- and Magnetoencephalographic Methods in Psychology 581
Eddie Harmon-Jones, David M. Amodio, Philip A. Gable, and Suzanne Dikker
Chapter 28. Event-Related Potentials 605
Steven J. Luck
Chapter 29. Functional Neuroimaging 631
Megan T. deBettencourt, Wilma A. Bainbridge, Monica D. Rosenberg
Chapter 30. Noninvasive Stimulation of the Cerebral Cortex 655
Dennis J. L. G. Schutter
Chapter 31. Combined Neuroimaging Methods 673
Marius Moisa and Christian C. Ruff
Chapter 32. Neuroimaging Analysis Methods 697
Yanyu Xiong and Sharlene D. Newman

Part IV. Psychometrics 721

Chapter 33. Reliability 723
Sean P. Lane, Elizabeth N. Aslinger, and Patrick E. Shrout
Chapter 34. Generalizability Theory 745
Xiaohong Gao and Deborah J. Harris
Chapter 35. Construct Validity 769
Kevin J. Grimm and Keith F. Widaman
Chapter 36. Item-Level Factor Analysis 793
Nisha C. Gottfredson, Brian D. Stucky, and A. T. Panter
Chapter 37. Item Response Theory 809
Steven P. Reise and Tyler M. Moore
Chapter 38. Measuring Test Performance With Signal Detection Theory Techniques 837
Teresa A. Treat and Richard J. Viken

Index 859

Copyright American Psychological Association

Contents

Volume 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological

Editorial Board xi
Contributors xiii

Part I. Qualitative Research Methods 1

Section 1. Overview of Qualitative Methods 3

Chapter 1. Developments in Qualitative Inquiry 5
Sarah Riley and Andrea LaMarre
Chapter 2. Metasynthesis of Qualitative Research 25
Sally Thorne

Section 2. Thematic Approaches 43

Chapter 3. Grounded Theory and Psychological Research 45
Robert Thornberg, Elaine Keane, and Malgorzata Wójcik
Chapter 4. Thematic Analysis 65
Virginia Braun and Victoria Clarke
Chapter 5. Phenomenological Methodology, Methods, and Procedures for Research in Psychology 83
Frederick J. Wertz

Section 3. Narrative and Language-Based Approaches 107

Chapter 6. Narrative Analysis 109
Javier Monforte and Brett Smith
Chapter 7. Ethnomethodology and Conversation Analysis 131
Paul ten Have
Chapter 8. Discourse Analysis and Discursive Psychology 147
Chris McVittie and Andy McKinlay

Copyright American Psychological Association

Section 4. Multilayered Approaches 169

Chapter 9. Ethnography in Psychological Research 171
Elizabeth Fein and Jonathan Yahalom
Chapter 10. Visual Research in Psychology 189
Paula Reavey, Jon Prosser, and Steven D. Brown
Chapter 11. Researching the Temporal 215
Karen Henwood and Fiona Shirani

Part II. Working Across Epistemologies, Methodologies, and Methods 233

Chapter 12. Mixed Methods Research in Psychology 235
Timothy C. Guetterman and Analay Perez
Chapter 13. The "Cases Within Trials" (CWT) Method: An Example of a Mixed-Methods Research Design 257
Daniel B. Fishman
Chapter 14. Researching With American Indian and Alaska Native Communities: Pursuing Partnerships for Psychological Inquiry in Service to Indigenous Futurity 285
Joseph P. Gone
Chapter 15. Participatory Action Research as Movement Toward Radical Relationality, Epistemic Justice, and Transformative Intervention: A Multivocal Reflection 305
Urmitapa Dutta, Jesica Siham Fernández, Anne Galletta, and Regina Day Langhout

Part III. Sampling Across People and Time 329

Chapter 16. Introduction to Survey Sampling 331
Roger Tourangeau and Ting Yan
Chapter 17. Epidemiology 359
Rumi Kato Price and Heidi H. Tastet
Chapter 18. Collecting Longitudinal Data: Present Issues and Future Challenges 385
Simran K. Johal, Rohit Batra, and Emilio Ferrer
Chapter 19. Using the Internet to Collect Data 409
Ulf-Dietrich Reips

Part IV. Building and Testing Models 433

Chapter 20. Statistical Mediation Analysis in Psychological Research 435
David P. MacKinnon, JeeWon Cheong, Angela G. Pirlott, and Heather L. Smyth
Chapter 21. Structural Equation Modeling With Latent Variables 459
Rick H. Hoyle and Nisha C. Gottfredson
Chapter 22. Mathematical Psychology 491
Parker Smith, Yanjun Liu, James T. Townsend, and Trisha Van Zandt
Chapter 23. Computational Modeling 515
Adele Diederich
Chapter 24. Fundamentals of Bootstrapping and Monte Carlo Methods 537
William Howard Beasley, Patrick O'Keefe, and Joseph Lee Rodgers

Copyright American Psychological Association

Chapter 25. Designing Simulation Studies 559
Xitao Fan
Chapter 26. Bayesian Modeling for Psychologists: An Applied Approach 579
Fred M. Feinberg and Richard Gonzalez

Part V. Designs Involving Experimental Manipulations 605

Section 1. Designs With Different Participant Assignment Mechanisms 607

Chapter 27. Randomized Designs in Psychological Research 609
Larry Christensen, Lisa A. Turner, and R. Burke Johnson
Chapter 28. Nonequivalent Comparison Group Designs 629
Henry May and Zachary K. Collier
Chapter 29. Regression Discontinuity Designs 653
Charles S. Reichardt and Gary T. Henry

Section 2. Experimental Manipulations in Applied Settings 671

Chapter 30. Treatment Validity for Intervention Studies 673
Dianne L. Chambless and Steven D. Hollon
Chapter 31. Translational Research 703
Michael T. Bardo, Christopher Cappelli, and Mary Ann Pentz
Chapter 32. Program Evaluation: Outcomes and Costs of Putting Psychology to Work 725
Brian T. Yates

Part VI. Quantitative Research Designs Involving Single Participants or Units 745

Chapter 33. Single-Case Experimental Design 747
John M. Ferron, Megan Kirby, and Lodi Lipien
Chapter 34. Time Series Designs 763
Bradley J. Bartos, Richard McCleary, and David McDowall

Part VII. Designs in Neuropsychology and Biological Psychology 785

Section 1. Neuropsychology 787

Chapter 35. Case Studies in Neuropsychology 789
Randi C. Martin, Simon Fischer-Baum, and Corinne M. Pettigrew
Chapter 36. Group Studies in Experimental Neuropsychology 807
Avinash R. Vaidya, Maia S. Pujara, and Lesley K. Fellows

Section 2. Genetic Methods in Psychology 825

Chapter 37. Genetic Methods in Psychology 827
Terrell A. Hicks, Daniel Bustamante, Karestan C. Koenen, Nicole R. Nugent, and Ananda B. Amstadter
Chapter 38. Human Genetic Epidemiology 849
Floris Huider, Lannie Ligthart, Yuri Milaneschi, Brenda W. J. H. Penninx, and Dorret I. Boomsma

Index 873

Copyright American Psychological Association

Contents

Volume 3. Data Analysis and Research Publication

Editorial Board xi
Contributors xiii

Part I. Quantitative Data Analysis 1

Section 1. Preparing Data for Analysis 3

Chapter 1. Methods for Dealing With Bad Data and Inadequate Models: Distributions, Linear Models, and Beyond 5
Rand R. Wilcox and Guillaume A. Rousselet
Chapter 2. Maximum Likelihood and Multiple Imputation Missing Data Handling: How They Work, and How to Make Them Work in Practice 27
Timothy Hayes and Craig K. Enders
Chapter 3. Exploratory Data Analysis. 53
Paul F. Velleman and David C. Hoaglin

Section 2. Describing Data 75

Chapter 4. Graphic Displays of Data. 77
Leland Wilkinson
Chapter 5. Estimating and Visualizing Interactions in Moderated Multiple Regression 111
Connor J. McCabe and Kevin M. King
Chapter 6. Effect Size Estimation 129
Michael Borenstein
Chapter 7. Measures of Clinically Significant Change 147
Russell J. Bailey, Benjamin M. Ogles, and Michael J. Lambert

Copyright American Psychological Association

Section 3. Methods With Single Outcomes 167

Chapter 8. Analysis of Variance and the General Linear Model 169
James Jaccard and Ai Bo
Chapter 9. Generalized Linear Models 201
David Rindskopf
Chapter 10. Multilevel Modeling for Psychologists 219
John B. Nezlek

Section 4. Methods With Outcomes Measured Over Time 243

Chapter 11. Longitudinal Data Analysis 245
Andrew K. Littlefield
Chapter 12. Event History Analysis 269
Fetene B. Tekle and Jeroen K. Vermunt
Chapter 13. Latent State–Trait Models 297
Rolf Steyer, Christian Geiser, and Christiane Loßnitzer
Chapter 14. Latent Variable Modeling of Continuous Growth 317
David A. Cole, Jeffrey A. Ciesla, and Qimin Liu
Chapter 15. Dynamical Systems and Differential Equation Models of Change 337
Steven M. Boker and Robert G. Moulder
Chapter 16. A Multivariate Growth Curve Model for Three-Level Data 351
Patrick J. Curran, Chris L. Strauss, Ethan M. McCormick, and James S. McGinley

Section 5. Multivariate Methods 377

Chapter 17. Exploratory Factor Analysis and Confirmatory Factor Analysis 379
Keith F. Widaman and Jonathan Lee Helm
Chapter 18. Latent Class and Latent Profile Models 411
Brian P. Flaherty, Liying Wang, and Cara J. Kiff
Chapter 19. Decision Trees and Ensemble Methods in the Behavioral Sciences 429
Kevin J. Grimm, Ross Jacobucci, and John J. McArdle

Section 6. Dyadic and Social Network Data 449

Chapter 20. Using the Social Relations Model to Understand Interpersonal Perception and Behavior 451
P. Niels Christensen, Deborah A. Kashy, and Katelin E. Leahy
Chapter 21. Dyadic Data Analysis 465
Richard Gonzalez and Dale Griffin

Copyright American Psychological Association

Section 7. Using Data Collected by Others 479

Chapter 22. The Data of Others: New and Old Faces of Archival Research 481
Sophie Pychlau and David T. Wagner
Chapter 23. Social Network Analysis in Psychology: Recent Breakthroughs in Methods and Theories. 501
Wei Wang, Tobias Stark, James D. Westaby, Adam K. Parr, and Daniel A. Newman
Chapter 24. Meta-Analysis 539
Jeffrey C. Valentine, Therese D. Pigott, and Joseph Morris

Part II. Publishing and the Publication Process 561

Chapter 25. Research Data Management and Sharing 563
Katherine G. Akers and John A. Borghi
Chapter 26. Questionable Practices in Statistical Analysis 579
Rex B. Kline
Chapter 27. Ethical Issues in Manuscript Preparation and Authorship 597
Jennifer Crocker

Index 617

Copyright American Psychological Association

About the Editor-in-Chief

Harris Cooper, **PhD**, is the Hugo L. Blomquist Professor, Emeritus, in the Department of Psychology and Neuroscience at Duke University. His research interests follow two paths. The first concerns research synthesis and research methodology. His book *Research Synthesis and Meta-Analysis: A Step-by-Step Approach* (2017) is in its fifth edition. He is the coeditor of the *Handbook of Research Synthesis and Meta-Analysis* (3rd ed.; 2019). In 2007, Dr. Cooper was the recipient of the Frederick Mosteller Award for contributions to research synthesis methodology given by the International Campbell Collaboration. In 2008, he received the Ingram Olkin Award for distinguished lifetime contribution to research synthesis from the Society for Research Synthesis Methodology. Dr. Cooper also studies the application of social and developmental psychology to education policy. In particular, he studies the relationship between time and learning.

Dr. Cooper chaired the first American Psychological Association (APA) committee that developed guidelines for information about research that should be included in manuscripts submitted to APA journals. In 2011, he published a book on the topic, *Reporting Research in Psychology: How to Meet Journal Article Reporting Standards*. In 2020, Dr. Copper published a revised second edition of the book as *Reporting Quantitative Research in Psychology: How to Meet APA Style Journal Article Reporting Standards*.

Dr. Cooper currently serves as the editor of *American Psychologist*, the flagship journal of APA. He served as editor for the *Psychological Bulletin* from 2003 through mid-2009. *Psychological Bulletin* is in the top five social science journals in total citations and impact factor. He was the chair of the APA Council of Editors in 2006 and was a member of the committee that revised the APA *Publication Manual* (2010). In 2012, Dr. Cooper became the inaugural coeditor of the *Archives of Scientific Psychology*, APA's first open methods, collaborative data sharing, open access journal. He remained as editor until 2015.

From 2009 to 2015, Dr. Cooper served as the chief editorial advisor for APA's journal publishing program. In this role, he served as a resource to the editors of APA's 70+ journals as well as the mediator of disputes between editors and authors and between authors and authors. Dr. Cooper's book *Ethical Choices in Research: Managing Data, Writing Reports, and Publishing Results in the Social Sciences* (2016) draws from the experience. The book goes beyond the proper treatment of human research subjects to examine frequently neglected ethical issues that arise after data have been collected.

Dr. Cooper served as the chair of the Department of Psychology and Neuroscience at Duke University from 2009 to 2014. He also served as chair of the Department of Psychological Sciences at the University of Missouri and director of Duke University's Program in Education. From 2017 to 2018 he served as the dean of social sciences at Duke.

Copyright American Psychological Association

About the Associate Editors

Marc N. Coutanche, **PhD**, is an associate professor of psychology and research scientist in the Learning Research and Development Center at the University of Pittsburgh. Dr. Coutanche directs a program of cognitive neuroscience research that uses brain imaging and behavioral studies to understand human learning, memory, and perception, particularly how the human brain transforms perceptual experience into knowledge. Addressing this question involves understanding the neural basis for successful memory encoding, the role of sleep in memory consolidation, how word learning is influenced by context, and more. In parallel, he develops and tests new computational techniques to identify and understand the neural information present within neuroimaging data. His work has been funded by the National Institutes of Health, National Science Foundation, American Psychological Foundation, and other organizations. He has published in a variety of journals that include *Trends in Cognitive Sciences; Cerebral Cortex; The Journal of Experimental Psychology: General; Memory & Cognition*; and *Perspectives on Psychological Science*. Dr. Coutanche received his undergraduate training in experimental psychology at Oxford University, earned a master's degree and PhD from the University of Pennsylvania, and conducted postdoctoral training at Yale University. He received a Howard Hughes Medical Institute International Student Research Fellowship and a Ruth L. Kirschstein Postdoctoral National Research Service Award, and was named a 2019 Rising Star by the Association for Psychological Science.

Linda M. McMullen, **PhD**, is professor emerita of psychology at the University of Saskatchewan (Canada). Over her career, she has contributed to the development of qualitative inquiry in psychology through teaching, curriculum development, and pedagogical scholarship; original research; and service to the qualitative research community.

Dr. McMullen introduced qualitative inquiry into both the graduate and undergraduate curricula in her home department, taught courses at both levels for many years, and has published articles, coedited special issues, and written a book (*Essentials of Discursive Psychology*) that is part of the American Psychological Association's (APA's) series on qualitative methodologies. She is also coauthor (with Frederick J. Wertz, Kathy Charmaz, Ruthellen Josselson, Rosemarie Anderson, and Emalinda McSpadden) of *Five Ways of Doing Qualitative Analysis: Phenomenological Psychology, Grounded Theory, Discourse Analysis, Narrative Research, and Intuitive Inquiry*, which shows how

Copyright American Psychological Association

the same set of data is analyzed using each of five leading qualitative methodologies. For the past couple of decades, her research has focused on discursive analyses of service providers' and service users' accounts of depression and the use of antidepressants. She is coeditor (with Janet M. Stoppard) of *Situating Sadness: Women and Depression in Social Context*.

Dr. McMullen has been engaged with building the Society for Qualitative Inquiry in Psychology (SQIP; a section of Division 5 [Quantitative and Qualitative Methods] of APA) into a vibrant scholarly society since its earliest days. She was a member of the Executive Committee from 2013 to 2021, served as its president in 2015–2016, and was elected SQIP Section Representative to the Division in 2018 for a 3-year term. In this latter capacity, she headed the multiyear process of developing an organizational structure for SQIP, codifying it in a set of bylaws, and harmonizing the SQIP bylaws with those of Division 5.

While working as a university professor, Dr. McMullen took on many leadership roles, including director of clinical training for the graduate program in clinical psychology at the University of Saskatchewan (1988–1994; 1995–1997), head of the Department of Psychology (1997–2002; 2003–2006), university leader for the Social Sciences and Humanities Research Council of Canada (2011–2013), and acting vice-dean for social sciences (2012–2014). She served as the elected faculty member on the university's Board of Governors from 2004 to 2010.

Dr. McMullen's contributions have been recognized by Division 5 of the APA (2021 Distinguished Contributions to Qualitative Inquiry Award), the Canadian Psychological Association (2003, Fellow; 2012 Distinguished Member, Section for Women and Psychology), and the Saskatchewan Psychological Association (1994 Award for Outstanding and Longstanding Service to the Profession).

A. T. Panter, **PhD**, is the senior associate dean for undergraduate education and a professor of psychology in the L. L. Thurstone Psychometric Laboratory at University of North Carolina at Chapel Hill. She received her BA from Wellesley College in 1985 and her PhD from New York University in 1989. She is past president of the American Psychological Association's (APA's) Division on Quantitative and Qualitative Psychology. As a quantitative psychologist, she develops instruments, research designs, and data-analytic strategies for applied research questions in higher education, personality, and health. Dr. Panter serves as a program evaluator for UNC's Chancellor's Science Scholars Program, a multisite adaptation of the successful Meyerhoff Program. She was also principal investigator for The Finish Line Project, a $3 million First in the World grant from the U.S. Department of Education that systematically investigated new supports and academic initiatives, especially for first-generation college students. Her books include the *APA Dictionary of Statistics and Research Methods* (2013), *The APA Handbook of Research Methods in Psychology* (1st ed.; 2012), the *Handbook of Ethics in Quantitative Methodology* (2011), *The SAGE Handbook of Methods in Social Psychology* (2004), and volumes on program evaluation for HIV/AIDS multisite projects.

At the undergraduate level, she teaches statistics, research methods, and a first-year seminar on communicating research results to others ("Talking About Numbers"). At the doctoral level, she teaches courses in research design, classical and modern approaches to instrument/survey design, and test theory and multivariate methods. Dr. Panter has received numerous awards for her teaching and mentoring, including the

Copyright American Psychological Association

Tanner Award, the J. Carlyle Sitterson Award, a Bowman and Gordon Gray Distinguished Professorship, and APA's Jacob Cohen Award for Distinguished Teaching and Mentoring. She is an APA Fellow, a member of the Graduate Record Examination Advisory Board, a former member of a Social Security Administration advisory panel related to disability determination, and a member of APA's Committee on Psychological Testing and Assessment. She regularly provides services for federal agencies, national advisory panels, and editorial boards.

Dr. Panter has been a member of the university faculty since 1989. As senior associate dean, she oversees all of the College's programs for undergraduate education: academic advising, undergraduate research, student success and academic counseling (including Learning Center, the Writing Center, Peer Mentoring, Carolina Firsts, Summer Bridge), undergraduate curricula (including the general education curriculum), instructional innovation, research and evaluation, Robertson Scholars Leadership Program, and Honors Carolina (including distinguished scholarships). Among her active work on campus, she helped design and implement the IDEAS in Action general education undergraduate curriculum, developed student learning outcomes for the university system, implemented a new holistic Thrive academic advising approach, address legal mandates related to the use of race/ethnicity in undergraduate admissions decisions, developed visualizations for key university data, and increased the number and type of high impact academic experiences for all undergraduate students.

David Rindskopf, **PhD**, is a Distinguished Professor at the City University of New York Graduate Center, specializing in research methodology and statistics. His main interests are in Bayesian statistics, causal inference, categorical data analysis, meta-analysis, and latent variable models. He is a Fellow of the American Statistical Association and the American Educational Research Association and is past president of the Society of Multivariate Experimental Psychology and the New York Chapter of the American Statistical Association.

Kenneth J. Sher, **PhD**, is Chancellor's Professor and Curators' Distinguished Professor of Psychological Sciences, Emeritus, at the University of Missouri. He received his undergraduate degree from Antioch College (1975), his PhD in clinical psychology from Indiana University (1980), and his clinical internship training at Brown University (1981). His primary areas of research focus on etiological processes in the development of alcohol dependence, factors that affect the course of drinking and alcohol use disorders throughout adulthood, longitudinal research methodology, psychiatric comorbidity, and nosology. At the University of Missouri, he has directed the predoctoral and post-doctoral training program in alcohol studies, and his research has been continually funded by the National Institute on Alcohol Abuse and Alcoholism for more than 35 years. Dr. Sher is a Fellow of the American Association for the Advancement of Science, Association for Psychological Science, and American Psychological Association (APA), and his research contributions have been recognized by professional societies, including the Research Society on Alcoholism (where he was awarded the Young Investigator Award, the Distinguished Researcher Award, and the G. Alan Marlatt Mentoring Award) and APA (Distinguished Scientific Contribution Award and the Distinguished Career Contributions to Education and Training from the Division on Addictions and a presidential citation), as well as by the University of Missouri

Copyright American Psychological Association

(where he received the Chancellor's Award for Research and Creativity, the President's Award for Research and Creativity, and the SEC Faculty Achievement Award/Professor of the Year), the National Institutes of Health (a MERIT Award, a Senior Scientist and Mentoring Award, and a Mark Keller Lectureship), and Indiana University (Richard C. Atkinson Lifetime Achievement Award). Throughout his career he has been heavily involved in service to professional societies (e.g., he served as president of the Research Society on Alcoholism and served on the APA's Council of Representatives, Board of Scientific Affairs, and Policy and Planning Board, chairing the latter two of these bodies). Dr. Sher also has a long history of service to scholarly publications, serving as an associate/field editor for both disciplinary journals (*Clinical Psychological Science, Journal of Abnormal Psychology*, and *Psychological Bulletin*) and specialty journals (*Journal of Studies on Alcohol and Drugs* and *Alcoholism: Clinical and Experimental Research*). He recently served as acting editor for *Clinical Psychological Science*. His current work is focused on attempting to define the "core" of addiction and developing methods to improve the diagnosis of alcohol use disorders and other psychological disorders.

Copyright American Psychological Association

Introduction: Objectives of Psychological Research and Their Relations to Research Methods

This is the second edition of *APA Handbook of Research Methods in Psychology*. It is the first handbook in the APA series to undergo revision. This seems only appropriate given the handbook's broad overview of research methods used throughout scientific psychology. Not surprisingly, research methods have evolved over the past 10 years since the first edition, and there was much new material to cover. Of 103 chapters in the revised handbook, 27 cover new topics or revisit topics but have been rewritten, 75 have undergone revision, many extensively, some with smaller changes and updated references. One chapter has been reprinted.

METHODS OF KNOWING

The American philosopher Charles Peirce (1839–1914) claimed that we use five different ways to decide what we believe is true about our world (Feibleman, 1969). First, we believe some things are true because authorities we trust tell us so. Sometimes, we know these authorities personally, such as our parents and teachers. Sometimes, they are very distant from us, such as the writers of ancient religious texts. Other times, authorities are less far removed but still not personally known, for example, the authors in a handbook on research methods.

Second, we know things are true because we have heard them repeated many times. Peirce called this the *method of tenacity*, or the *a priori method*. Here, something is believed because it has always been believed (longevity) or because we have heard it repeated many times. We could include in this method the commonsense adages with which we are all familiar, such as "birds of a feather flock together" or "a stitch in time saves nine."

Third, we observe or experience things ourselves and our senses tell us they are true. The sun warms things up, for example.

Fourth, we know that some things are true because they can be logically derived; they are the product of rational analysis. Without getting into the formalities of logical

My sincere thanks go to the five associate editors of this handbook, Marc N. Coutanche, Linda M. McMullen, A. T. Panter, David Rindskopf, and Kenneth J. Sher. The latter three also served as editors on the first edition. They deserve equal billing with me as editors. They also provided feedback on this introduction, although any mistakes remain my doing. A special note of thanks goes to Kristen Knight, APA project editor, for her diligence and organizational efforts.

Copyright American Psychological Association

deduction, if a premise known to be true tells us that "all males have an Adam's apple," and we observe a men's intercollegiate fencing match, then logic dictates we believe that these fencers have Adam's apples under their masks.

The problem with each of these first four methods of knowing is that they are fallible. Two trusted authorities can disagree, suggesting that one (at least) must be wrong. Tenacious beliefs can lead us astray because conditions change over time, or what seems like common sense is not so sensible after all (remember, in addition to birds of a feather flocking together "opposites attract," and although a stitch in time may be frugal, "haste makes waste"). Our senses can deceive us, for example, through the application of different frames of reference, as demonstrated by optical illusions. Are the soccer balls in Figure 1 the same size?

Finally, a logical deduction is based on the validity of the premises, which may be wrong. Or, the logic itself may be faulty even though the premises are true, as when we affirm the consequent ("All male fencers have an Adam's apple," and "Spencer has an Adam's apple"; therefore, "Spencer is a fencer").

Peirce's final method of knowing was the scientific method. We can think of the scientific method as a combination of observation and rational analysis, or observation using a set of logical rules that should lead to veridical conclusions about the world. Peirce expected the scientific method to lead to knowledge that was exactly the same for every person, uninfluenced by idiosyncratic frames of reference. He held out the hope for truly objective knowledge.

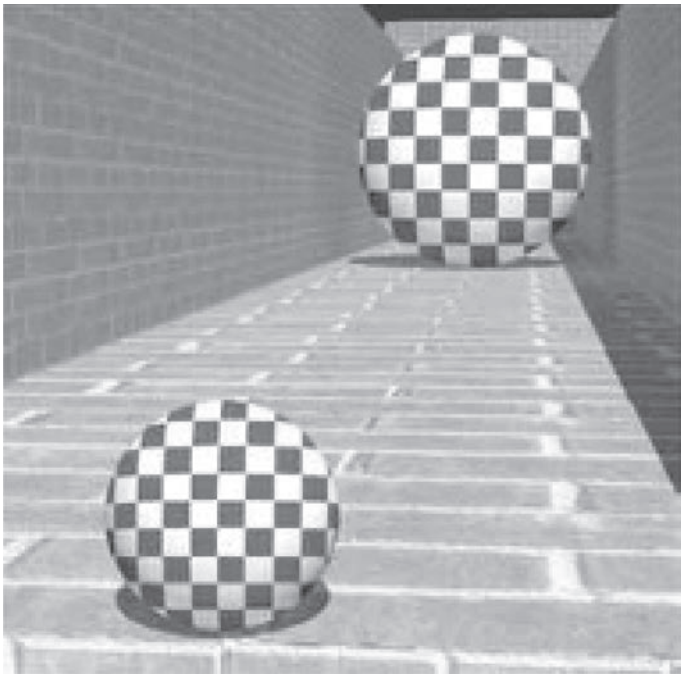


FIGURE 1. Optical illusion involving two balls of the same size. From "The Representation of Perceived Angular Size in Human Primary Visual Cortex," by S. O. Murray, H. Boyaci, and D. Kersten, 2006, *Nature Neuroscience 9*, pp. 429–434 (https://doi.org/10.1038/nn1641). Copyright 2006 by Springer Nature. Adapted with permission.

Copyright American Psychological Association

THE FLAWED BUT SELF-CRITICAL NATURE OF SCIENTIFIC INVESTIGATION

Peirce's grand vision for science, especially when applied to the social and behavioral sciences, is viewed by many in the 21st century as naive, at best. Vigorous debate prevails about whether knowledge is ever attainable without being infused with the theory of the observer (Godfrey-Smith, 2021; Kuhn, 1996). And studies that turn the methods of science on itself (the sociology of science) suggest that even when the search for objective knowledge is a desirable (and an obtainable) goal, the scientific process, in practice, is still replete with appeals to authority, idiosyncratic observation, and failures of rationality (Merton, 1957, 1979).

Perhaps it is best, then, to thank Peirce for having pointed out potential flaws in the ways that we acquire knowledge. And, although his belief in the existence of objective truth is open to debate, he argued that the application of rationally derived rules to observation was a self-correcting system. Over time, he asserted, by putting the claims of our authorities, observations, tenacious beliefs, and logical deductions to repeated empirical testing (with further observation and rational analysis), our erroneous beliefs would be replaced by *truer* ones.

The view of rational analysis as self-correcting captures only a secondary characteristic of the scientific method, in the case of social science, the scientific study of thought, feeling, and behavior. At the heart of the scientific ethos is the notion of critical analysis. That is, a scientific posture requires that we be skeptical of *any* truth claim, no matter from where it comes, even from science itself. Scientists test ideas in multiple ways from multiple perspectives with the *failure to disprove* after rigorous testing as the goal of their efforts.

Let's use Figure 1 to construct an example. In an ideal world, a scientist says,

> My observation suggests that the upper ball is larger than the lower ball. Now, let's try to prove that my observation is wrong.[1] As a public demonstration, I will use my thumb and forefinger to measure the diameter of the lower ball; then, holding my hand steady, I will move my measurement to the upper ball.

The measurement would suggest that the balls were roughly equal in diameter. Another scientist might then point out that my thumb and forefinger could not be held perfectly steady and suggest a ruler be used. The ruler would still suggest that the balls were equal in diameter. The process might continue with progressively more precise and careful measurement. Eventually, the observing scientists would come to agree that the hypothesis that the balls were the same size could not be rejected, or that evidence they were different sizes had failed repeated tests.

Next, the scientists would turn their attention to discovering why their eyes had deceived them. Then, as plausible evidence accumulated about *why* the reliance on simple observation was flawed (perhaps gathered by conducting experiments that manipulate the angle of the converging walls or the shadows in Figure 1), confidence in the conclusion about the state of nature, or the laws of visual perception, will grow.

[1]To be more precise, in the tradition of null hypothesis testing the scientist might say, "I propose an alternative to the hypothesis that the two balls have equal diameter. I propose that the upper ball is larger. Now, let's try to reject the equal diameter hypothesis."

Copyright American Psychological Association

A self-critical posture requires that truth claims never be believed with absolute certainty, only with greater or lesser certainty. This is what is most unique and, I think, exciting about the scientific posture. It is also one tenet that binds the chapters in this handbook. All of the chapter authors would agree that psychological scientists must take a critical stance toward what they call "knowledge."

SCIENCE AND DEMOCRACY

There is much to admire in this self-critical stance to knowledge acquisition. Importantly, in the 10 years since the first edition of this handbook, the need for critical thinking has mushroomed into a serious societal concern. Polarized thinking that is immune to evidence has driven deep divides in the body politic (Rutjens et al., 2018). While the self-critical methods outlined herein are essential to the solution of so many social problems, science itself is under attack and needs defense.

Timothy Ferris (2010) claimed that science is inherently antiauthoritarian because of its skeptical stance. In fact, he claimed that science and liberal democracy, that is, a social system that values human rights and freedom of action, go hand in hand. One cannot flourish without the other. Ferris wrote,

> The very process of doing first-rate science—of making important discoveries rather than merely refining old ideas—depends on unfamiliar and unpopular ideas being freely promulgated, discussed, and in some instances accepted. The fact that millions of people today are open to new ideas and skeptical about political and intellectual authority is largely due to the rise of science. (p. 4)

So, I would add another attribute shared by the authors of chapters in this handbook: open-mindedness. At the same time scientists accept no truth claim uncritically they also turn no idea away prima facie, on its first appearance.

Maintaining this posture of "open-minded skepticism" is no easy task. Finding the proper balance between foregoing prejudgment and embracing doubt, while holding in abeyance our only-too-human desire for certainty (and for *our* certainties to be deemed the correct ones), is the scientist's principal challenge. And psychological scientists have a particularly difficult task. They must hold to open-minded skepticism while studying their own species in contexts in which they themselves act. Unlike the physicist who can remain relatively detached from the behavior of atoms, psychological scientists can have a personal stake in what they discover about human nature, the good and the bad. So, is open-minded skepticism impossible? Perhaps. Flawed in its execution? More often than we might desire. Worth the effort? Certainly.

In the pages of this handbook, you will find descriptions of many techniques that psychologists and others have developed to help them pursue a shared understanding of why humans think, feel, and behave the way they do. These are the tools that we use to conduct our rational, self-critical, and open-minded analyses.

THE HANDBOOK'S ORGANIZATION

Organizing the chapters of this handbook was a huge challenge. Psychology's methods defy simple categorization because of their cross-disciplinary (and subdisciplinary) heritages. Many methods presented are similar to one another on some dimensions but

Copyright American Psychological Association

far apart on others. So, deciding which dimensions to prioritize has nontrivial implications for where a method appears. In addition, prioritizing some dimensions over others can border on arbitrary and be based on oversimplified characterizations of any methodology's capacities for guiding discovery. Many methods can be used for more than one purpose. We have tried to put these "Swiss Army knives" of methodology in the toolbox compartment of their most frequent use. In addition, traditions of use within subdisciplines dictated that some methods appear close together, even if grouping them defied the logic of our dimensional analysis. And, some methods are so unique that they were hard to place anywhere. These methods are no less important because of their singularity; indeed, if they are the only way to answer a question, their uniqueness can make them especially valuable.

So, as you scan the table of contents and contemplate our choices for clustering and ordering the presentation of methods, I am certain that you will be perplexed by some of our choices. Other schemes could fit equally well, or better. Below, I try to capture the high-order dimensions that informed our placement of chapters, beginning with those that relate to the earliest decisions that a researcher makes when choosing methods.

FINDING THE METHOD THAT FITS THE QUESTION

There is an old joke in which a person is searching the ground beneath the halo of a streetlight.

> A stranger emerges from the dark and asks, "What are you looking for?"
>
> "My car keys," replies the searcher.
>
> The stranger sees nothing under the light and inquires, "Where did you lose them?" "Over there," says the searcher pointing down the street.
>
> The stranger asks, "So why are you looking here?"
>
> "Because this is where the light is," the searcher explains.

Clearly, this searcher's method does not fit the objective. Similarly, psychological researchers must choose methods that fit the research question that they want to answer, not the method that is available or that they know best. No matter how luminous a method is, if the questions it can answer do not correspond to the knowledge sought, the researcher will remain in the dark.

You could think of this handbook as a collection of streetlights. Each method contained herein is meant to help you shed light on thought, feeling, and behavior over a different expanse and from a different angle. As I alluded to, another frequent metaphor compares research methods with a toolbox. Here, methodology provides the hammers, screwdrivers, wrenches, and rulers that psychological researchers use when they ply their trade.

You will read repeatedly in the chapters that follow that your first task as a psychological researcher is to pick the method best suited to answer the question that motivates you. You will be told not to search where the light is or bang a nail with a screwdriver. Instead, you will learn to choose the method that best answers your question. The contributors hope that this handbook will expand the topics that you can illuminate and increase the size of your toolbox. We hope to provide you with new ways to answer old questions as well as to raise new questions, perhaps ones you did not realize could be asked.

Copyright American Psychological Association

At the broadest level, when choosing a method you make decisions about (a) what measurement and data collection techniques best capture the thoughts, feelings, and behaviors that interest you; (b) what research design best fits the question that you want to answer; and (c) what strategies for data analysis best match the characteristics of your measurements and design.

The simplest choice for organizing the presentation of material is the temporal sequence in which you will make these decisions. This is roughly what we have done. So, the earliest chapters in Volume 1, Part I, address the broadest questions related to research designs. These involve both (a) which research designs are most appropriate for which question and (b) how to think about the ethicality the research that address your question, and (c) how to conduct research with participants drawn from more diverse populations.

Next, the chapters in Volume 1, Part II, help you with the research planning process, including how to develop a testable hypothesis, find the pertinent literature, secure the resources you need, and choose measures and what people to study. Part III of the first volume describes the plethora of measurement techniques that psychologists most often use to collect data and how to determine whether the measurement techniques that you might choose are the best ones for your purpose. For this revised edition of the handbook, significant changes were made to the discussions of chronometric and psychophysical measures (Section 4) and the measures used in psychophysiology and neuroscience (Sections 5 and 6).

Part IV contains chapter looking at different ways to assess the trustworthiness of measures. These help you determine whether your measures can (or did) allow you find the answers you sought.

In Volume 2, Parts I through VI, the chapters return to issues of research design. They present for your consideration a panoply of options, further divided along more nuanced distinctions in their objectives (discussed in the following section Interpretive Inquiry, Description, and Causal Explanation).

Chapters on techniques for data analysis follow in Volume 3, Part I, again with special attention to the fit between design, measurement, and analysis. Finally, issues and choices you must consider when you write up your research to share with the community of psychologists are discussed in the handbook's concluding chapters, in Volume 3, Part II.

INTERPRETIVE INQUIRY, DESCRIPTION, AND CAUSAL EXPLANATION

To choose the research design that best fits your research question, you need to consider some questions about your research aims. Are you seeking to (a) undertake an interpretive inquiry, (b) provide a description of an event, or (c) develop a causal explanation for the event or relation that interests you? Are you hoping to discover how individuals change over time or what makes groups of individuals different from one another, on average? The handbook begins with three examinations (including this Introduction and Chapters 1 and 2) that help you understand the differences between these types of questions. Then, in Volume 2, the chapters map specific research designs onto specific research questions.

Interpretive Research

To choose between interpretive or descriptive research, you must also answer the following question: Do you want to uncover the impetus to thoughts and actions that exist for the actors themselves or do you have your own theory or perspective to guide

Copyright American Psychological Association

your data collection? Carla Willig (Chapter 1 in this volume) helps you decide whether your question naturally fits in the former category, suggesting an interpretive inquiry design. She suggests that qualitative research designs are most appropriate when the interpretation of an event by participants is your goal or when your question falls into one of these categories:

- What does something feel like?
- How is something experienced?
- How do people talk about something and with what consequences?
- How do people make sense of an experience?
- How do they construct its meaning? What does this allow them to do or not to do? To feel or not to feel?
- How does a particular event unfold? How do participants experience the event? What may be its consequences? For them and for others?

Willig makes the point that these approaches to psychological research are most appropriate when the researchers do not want to impose their own (or someone else's) theory or perspective on the thoughts, feelings, or actions of the people that they are studying. Rather, the researchers want to uncover the impetus to behavior that exists for the actors themselves. Cultural anthropologists refer to this as using an *emic* approach to describing behaviors and (conscious or unconscious) beliefs.

Qualitative designs (detailed in Volume 2, Part I) use a range of data, including spoken or written narratives from interviews and informal conversations; archival data contained in public records and private diaries; and visual data from photographs, film, and video. Although these data are often obtained from relatively few participants, large-scale, primarily quantitative, studies have increasingly employed a range of qualitative methods to explore a diverse range of questions. In recent years, the desirability of using multiple methods in research has become increasingly evident and is being adopted more often. Volume 2, Part II contains new chapters that address the use of multiple methods for collecting both qualitative and quantitative evidence in the same study.[2]

It is also possible to take an *etic* approach to research, or to provide a description of an event. Here, the researchers' theories and beliefs are applied to the situations that they study. These forms of descriptive research often focus on a few specific characteristics of events or individuals chosen a priori by the researcher. Participants are then sometimes broadly sampled and they respond to questions developed by the researchers to investigate particular aspects of the person or situation. Similar to qualitative research, the researchers make no attempt to manipulate the participants' circumstance. Similar to quantitative research, the data collected will be in numerical form and examined through statistical procedures.

Causal Explanatory Research

If your answer to the first question was (c), that you were seeking a causal explanation for an event or relation, then you will be looking for a different type of research design. When an investigator is in search of a causal connection, research cannot be undertaken

[2]For an interesting take on the similarities and differences between quantitative and qualitative approaches, see Shweder (1996): "The true difference between the approaches is not over whether to count and measure but rather over what to count and measure, and over what one actually discovers by doing so" (p. 179).

Copyright American Psychological Association

without some theoretical underpinning. Sometimes the theory is explicit, sometimes implicit, but it is always there. Theory tells us what variables to focus on as potential causes and effects, or how to divide the world into meaningful chunks.

Even with a theory to guide you, however, coming to a firm conclusion that one event has caused another may be more problematic than it seems at first.[3] In fact, "How do we know a cause when we see one?" is a question that has vexed philosophers of science for centuries. To understand why, we need to digress into a bit of science history.

The Scottish philosopher David Hume (see *A Treatise on Human Nature*, 1739–1740/1978), set out the dilemma for us (and, some would say, led social scientists astray; see Maxwell, 2004). Hume argued that for something to be considered a cause: (a) the cause and the effect had to happen together, (b) the cause had to occur before the effect, and (c) there had to be a *necessary connection* between the two events. Agreeing on whether Hume's first two conditions prevail in a particular search for a causal relationship is relatively straightforward. A researcher needs to show that Events A and B co-occur more often than would be expected by chance (although chance, being the iffy thing it is, implies that we can never make this determination with absolute certainty). The temporal sequence of events is typically observable with a high degree of reliability (although sometimes events occur nearly simultaneously, and sometimes effects, in psychology at least, are caused by the anticipation of other events).

Hume's (1739–1740/1978) third condition presents the greatest challenge to researchers in search of causes. Hume argued that we can never know with perfect certainty that the event we are calling the cause was the necessary connection that produced the effect.

A thought experiment will clarify his claim. Suppose I placed my coffee cup on the edge of my desk. My elbow slid into a book that then knocked the cup to the floor. What caused the coffee cup to fall? If we were to ask a group of people who observed the event to independently identify the cause of the spill, we would be confronted with multiple nominations. Most observers would say "the book" or "your elbow," but the more playful in the group might nominate "a gust of air" or even perhaps "a poltergeist." Are they wrong? Can you prove it *conclusively*?

Hume (1739–1740/1978) asserted that events happen in an unending flow and that even designating where one event ends and the next begins is subjective, that is, requires a prior theory or perspective that is supplied by the observer.[4] Therefore, he claimed, whenever we identify a *cause*, it remains possible to argue for other causes in two ways. First, we can identify another event that takes place *between* the asserted cause and effect. So, if I claim my elbow caused the coffee cup to fall, you can counter-claim that it was the book that provided the necessary connection. Yet another observer (especially a physicist) could claim *gravity* was the cause. After all, without gravity, the coffee cup would have remained suspended in air. Events preceding my errant elbow might also be viable alternate nominations—it was my writing deadline that caused me to be sitting at my desk.

Second, Hume (1739–1740/1978) argued that causal systems are open to outside influences. That is, an outside event can enter the claimed causal sequence. The "gust of air" or "poltergeist" explanation for my coffee spill would be outside

[3]See Cooper (2007) for my first presentation of this material on Hume (1739–1740/1978).
[4]Note how this harkens back to the criticisms of Peirce.

Copyright American Psychological Association

"an elbow-book-gravity-spill" causal system. Could an invisible mischievous spirit have knocked the cup just a nanosecond before the book hit it?

If we accept Hume's (1739–1740/1978) argument that we never know causes with complete certainty, then how are we to proceed to answer the question "What events cause other events to happen?" Steiner et al. (Chapter 2 in this volume) present the three influential frameworks that most psychologists use—drawn from social science, statistics, and computer science—to decide how evidence can be mapped onto causal claims so as to make the claim more or less plausible. Each strategy has implications for how data will be collected and analyzed. Oversimplifying, these strategies suggest that if causal explanation is what you seek, you next must answer the question "Are you interested in (a) testing the implications of a causal model or (b) manipulating a possible cause to see whether it has the presumed effect?" Depending on where your interest lies, you would choose either a design that involves causal modeling or a manipulation of experimental conditions.

Causal modeling. Research designs that propose and test causal models and do not involve experimental manipulations could be categorized as *quantitative descriptive research*. Researchers who use modeling approaches play out the *implications* of different causal assumptions and, therefore, produce results that bear on the plausibility of causal relations (Neimark & Estes, 1967). This is especially true of researchers who build and test structural equation models. Steiner et al. (Chapter 2 in this volume) point out that causal models are often intended to provide an exhaustive description of a network of linkages, which when coupled with certain assumptions (this is where Hume's, 1739–1740/1978, ghost haunts us) imply causal relationships.

Typically (and again oversimplifying), the modeling approach begins when researchers propose a sequence of interconnections that they believe captures the underlying causes of thought, feeling, or behavior. Then, they use one of numerous approaches to see how well the model and the data fit. They examine the co-occurrence of events in a multivariate, temporally sequenced framework. So, for example, I might propose that the sales of research methods handbooks are caused by (a) the editor's level of compulsiveness, which affects (b) the level of expertise of those chosen to be associate editors. Next, the expertise of the associate editors affects (c) who is chosen to be chapter authors. Then, the expertise of the chapter authors influences (d) the audience's perception of the value of the book and, finally, (e) their decision to buy the handbook.

Rodgers (2010) viewed the ascendance of mathematical and statistical modeling in psychology as nothing less than a (quiet but much needed) epistemological revolution. For one thing, most causal models focus on multivariate and temporally sequenced descriptions of behavior. These are typically more complex, and more complete, than the descriptions that you might find tested in many other approaches to studying causal relationships. Also, users of the modeling approach are less interested in testing a model against a null hypothesis ("Is this model better than no model at all?") but rather against an alternative model ("Is this model better than another proposed model?"). So, my model of handbook sales might be compared with one in which the publisher's advertising budget was also included.

Experimental and quasi-experimental designs. When researchers control aspects of the experimental situation by the purposive manipulation of an event, typically they

Copyright American Psychological Association

do so to identify a cause-and-effect relation between one or a few presumed causes and one or a few effects rather than to investigate a complex and comprehensive model.

Designs are called *experimental* when they involve purposive manipulation of different conditions within the study. In this case, a study is conducted to isolate and draw a direct link between one event (the cause) and another (the effect). In studies that employ random assignment of participants to conditions, both the introduction of the event and who is exposed to it are controlled by the researchers (or other external agents), who then leave the assignment of conditions to chance. This approach is the best we have to ensure that on average the groups will not differ before the purposive manipulation. Therefore, we can be most confident (but not completely confident) that any differences between the conditions that we have created were caused by the manipulation, rather than preexisting differences between the participants in one condition from those in another.[5]

Designs with purposive manipulations can also be *quasi-experimental*. Here, the researchers (or some other external agents) control the introduction of the experimental manipulation but do not control precisely who may be exposed to it. In these designs, the researchers must find ways other than random assignment to equate participants in the various conditions so as to render less plausible the notion that preexisting differences between participants can explain any differences that they find on the outcome measures (the effects).

Individual-Change and Group-Difference Research

Before your quest for a research design can move from the general to the specific, you must also answer the question "Are you interested in understanding (a) how an individual (or more generally, a single unit) behaves or changes over time or (b) what makes one group different from another group, on average?" Let me give an illustration that demonstrates why this distinction is so important.

We know that sometimes learning, understanding, or cognitive change comes to a person as a sudden "aha" experience (Kounios & Beeman, 2009). The person gets it pretty much all at once, and a noticeable change in thinking becomes immediately evident and remains thereafter. Different people may have aha experiences after different numbers of exposures to an event or stimulus. For example, the top panel of Figure 2 displays the scores of six hypothetical people on a questionnaire regarding their perception of the need for energy conservation. The imaginary participants were exposed to 10 proconservation messages in the same sequence.[6] Each message highlighted a different reason for conserving energy—health threats from pollution, climate change, reduced costs, energy independence, and so on. On the basis of the data in this figure, the best description of how a person's perspective on the need for conservation changed would be to say that each person experienced a cognitive change on viewing a particular message, after which their perspective became noticeably more proconservation and remained so. But the message that precipitated the change was different for different people, and one person never changed at all.

[5]But we can never be completely certain that the characteristic of the manipulation that we claim is causal was the productive element because our experimental and comparison conditions can be viewed as differing on many characteristics. Hume's ghost again.

[6]You would not want to do this in a real experiment because it confounds the particular stimuli with the order. In this imaginary scenario, you must assume that order has no influence, so that differences between people are due to the message.

Copyright American Psychological Association

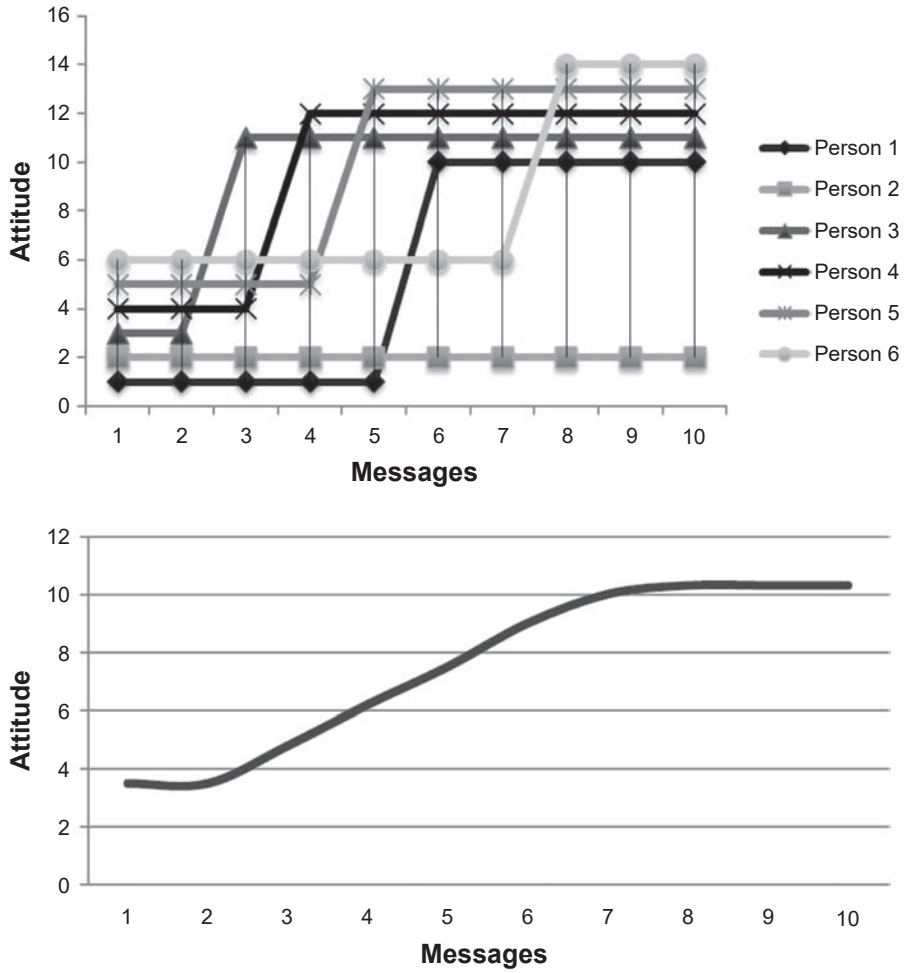


FIGURE 2. Hypothetical graphs of six individuals' change in attitude (top panel) and group-averaged attitudes of those six people (bottom panel) toward energy conservation after viewing 10 proconservation messages with different themes.

The bottom panel of Figure 2 provides a graph of how the same data look if averaged across the six people. A much different picture emerges. Looking at this graph, if you were to *assume* the group average effect accurately reflected what was happening to each individual person, you would say that after the second exposure, each person gradually changed their perspective on conservation after viewing each message. After eight exposures, no additional change took place. Clearly, this would be an incorrect characterization of the process occurring at the level of the individual person. That said, the group-averaged data could be used to describe how change occurred for the group as a whole, as a single unit. Thus, the correct interpretation of the group-averaged data would be to say that when a group was exposed to proconservation messages, the group average attitude changed gradually after the second exposure, but there was no additional change after eight exposures. This would be an accurate description of how the group behaved over exposures, but it would not adequately describe any single member within the group.

Whether you are interested in the individual or group-averaged effect depends on the context in which your question is being asked. Sometimes the group average does

Copyright American Psychological Association

represent the behavior of a single unit. So the bottom panel of Figure 2 is a description of, say, how one city's perspective might be described if the 10 messages were weekly programs aired on a local television channel. Then, the city becomes the unit, not the individuals who populate it. If your problem focuses on understanding how individual units change over time, then the top panel of Figure 2 provides you with six replications of the phenomena of interest. The bottom panel is irrelevant to your question. If your focus is on how a group average changes over time, the bottom panel provides you with one instance of this and the top panel is irrelevant.

Summary of Design Considerations

Figure 3 lays out the relations between the four questions about a research question. The figure needs to be read from both the top and the bottom to arrive at the designs in the middle. This is because two questions, those about interpretive, descriptive, or explanatory research (at the top) and single-unit versus differences-between-groups-of-units research (at the bottom), are independent of one another. Which of the two other questions you answer depends on how you answer the question about interpretation, description, or explanation. A theory-discovery or theory-specified approach is available to you once you have decided that your question is interpretive or descriptive. If your question is explanatory, it is theory driven by definition, but you must decide whether the question involves modeling the causal implications of the theory or estimating its causal effect via an experimental manipulation.

The handbook uses the answers to these questions to organize the chapters on research design found in Volume 2. Interpretive research designs that emphasize a qualitative

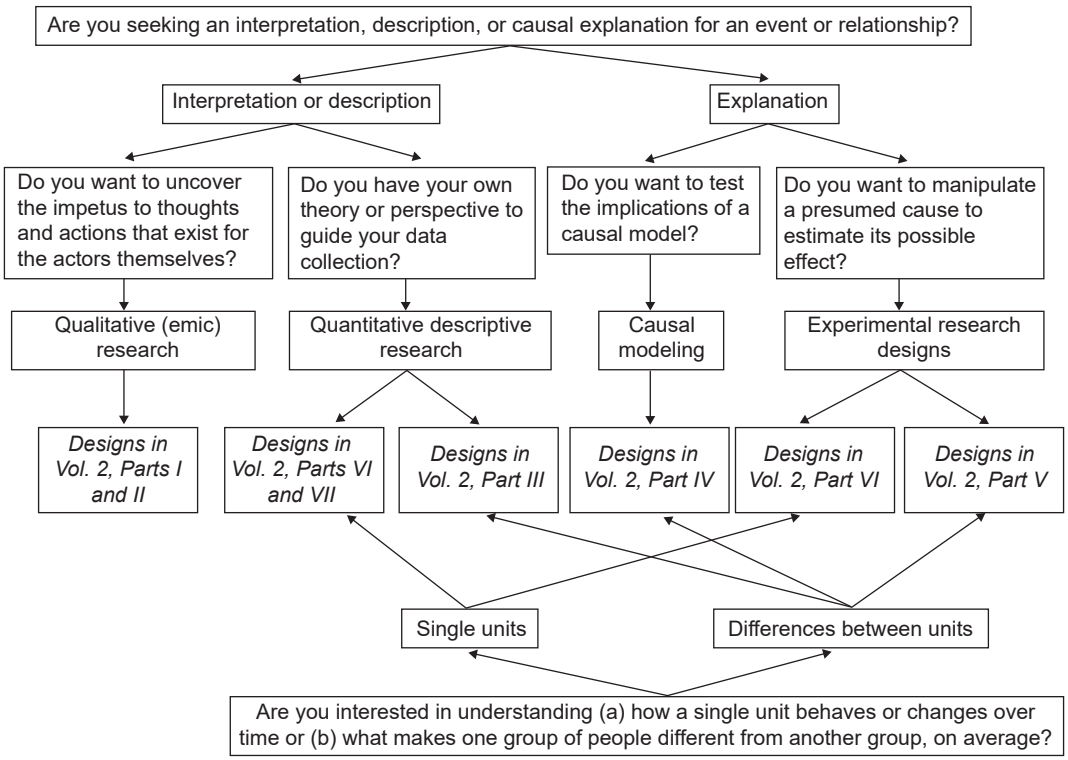


FIGURE 3. Relations between research questions, research designs, and the organization of parts in the *APA Handbook of Research Methods in Psychology*.

Copyright American Psychological Association

approach are detailed in Part I. Most of these designs also focus on questions that involve describing the current state or change in an individual unit of interest or a small sample of individuals. Part II looks at issues at the transition point between qualitative and quantitative research, as well as issues that arise in research with particular populations and settings that go beyond their simple participation in research.

Volume 2, Parts III through VII, introduce designs that emphasize an etic (or theory-specified), more quantitative approach to research. Volume 2, Part III, presents issues in sampling for quantitative studies. Although the techniques described herein will be of interest to all researchers, they would be of special interest to those who are conducting descriptive research. So, for example, if you are collecting data on the impact of growing up in a single-parent home, this section will assist you in planning your strategy for sampling respondents, help you consider issues you will encounter in collecting data from participants over time, and help you determine how you might use the internet. Volume 2, Part IV, focuses on designs that build and test the implications of causal models. You will find here approaches that differ in the type of data that they employ (sometimes even simulated data) and in the assumptions that are made as part of the data analysis (e.g., Bayesian modeling). Volume 2, Part V, focuses on research with experimental manipulations, in which participants are deliberately treated differently. Section 1 of Part V distinguishes these designs depending on how participants were assigned to their experimental conditions. Section 2 of Part V describes many of the unique problems that psychologists face when they conduct experimental research in applied settings.

In Volume 2, Part VI, designs are introduced that focus on theory-testing questions, rely heavily on quantification, and are used to study change in individual units. These designs all require multiple measurements of the same dependent variable(s) over a period of time. They can be used to study change that is either (a) naturally occurring, for example, as when a researcher wants to describe how a person's cognitive abilities change as they age, or (b) purposively manipulated, as when a researcher examines an older adult's cognitive ability before and after an intervention that is meant to improve memory.

The designs in Volume 2, Part VII, are labeled "Neuropsychology" and "Genetic Methods in Psychology." Here you will find designs for theory-driven research that derive largely from the more biological end of psychology's family tree.

MEASUREMENT METHODS AND PSYCHOMETRICS

After you have a good grasp of your research question and the general design of your study, you need to choose the means of measuring the variables of interest to you. You must answer the question "What measures best capture the variables of interest?" Volume 1, Parts III and IV of the handbook help you consider your choices and pick the measure that best answers your question.

Units of Analysis

As a science matures, it adopts, adapts, and invents new techniques for looking at the world. Certainly, the ruler that we used to examine the optical illusion in Figure 1 was not invented for that purpose alone. Rather, we realized its relevance to our problem and commandeered it for our cause. And, as you look at the contents of Volume 1,

Copyright American Psychological Association

Parts III and IV, it will be obvious to you that this handbook describes an enormous array of rulers. Some of these rulers were invented by psychologists, but many were invented for other purposes to study phenomena of interest in other disciplines.

It is possible to think of the sciences as falling along a continuum that distinguishes them according to the size of the things that they study or their unit of analysis or investigation. So chemists, generally speaking, study things that are physically smaller than the things studied by biologists, whose units of study often are smaller than those studied by psychologists, whose units often are smaller than those studied by sociologists. Of course, the overlap in topics of interest is great, so at the margins the distinction between disciplines breaks down; it becomes difficult, if not impossible, to identify where one scientific discipline ends and the next begins. A psychologist who studies group identity is more likely to share an intellectual heritage with many sociologists than with a psychologist who studies the role of neurotransmitters in depression, whose work may be more akin to that of a neurobiologist.

Along with a blurring at the margins of disciplines comes the transfer of measurements and methods between disciplines. Not surprisingly, then, in this handbook, you will find measurement techniques (as well as research designs and statistical techniques) with histories that locate their roots in numerous fields of study, including economics, political science, sociology, anthropology, neurobiology, and genetics. This is a good thing for our discipline. Psychologists have come to recognize that a complete picture of any phenomenon requires that it be examined through multiple methods, applying multiple rulers. To fully understand schizophrenia, for example, psychological scientists might need to survey its prevalence in a population, examine family dynamics, observe, interview, and test individuals with the disorder, conduct brain scans, and map genes.

Because of psychology's interdisciplinary range, the array of methods covered in this handbook is daunting. But the variety of methods that psychologists use is indicative of our discipline's strength and vitality. The authors of the handbook chapters are motivated by a search for answers, no parochialism here. They share the belief that their own method of choice cannot develop a complete picture of the world or, really, any discrete phenomenon in it. Rather, each method supplies a small piece of the puzzle. It is only when the puzzle pieces are fit together that a complete picture emerges.

Volume 1, Part III, of the handbook offers many different techniques of measurement. The sections are roughly organized according to the size of their unit of analysis. It begins with the largest units and proceeds to the smallest. So, Section 1 presents techniques that measure peoples' overt individual behavior, which are typically available for others to view. Sections 2 and 3 largely describe measures for which people provide verbal or written data about what they are thinking, what they are feeling, or how they behave. Sections 4, 5, and 6 reduce the unit of analysis even further, to psychophysical and psychophysiological measures and then to measures that are biological in nature.

The chapters in Volume 1, Part IV, help you answer a second question about your measures: "How well does your chosen measure represent the variable that interests you?" This question again requires you to consider fit, but now between a concept, or latent variable, and the means that are used to measure it. Put simply, the variables involved in psychological research need to be defined in two ways, conceptually and operationally. *Conceptual* definitions describe qualities of the variable that are independent of time and space and can be used to distinguish events that are and are not instances of the concept. For example, a conceptual definition of *aggression* might be "behavior intended

Copyright American Psychological Association

to cause pain or harm." Conceptual definitions can differ in breadth, that is, in the number of events that they capture. So, if the terms *pain* and *harm* are interpreted broadly, then *aggression* could include verbal as well as physical acts.

To relate concepts to concrete events, a variable must also be operationally defined. An *operational* definition is a description of the observable characteristics that allows us to determine whether a behavior or event represents an occurrence of the conceptual variable. So, an operational definition of *aggression* might include "shouting, or vocalizations above a specified decibel level" if verbal aggression is included but not so if only physical harm is included. The chapters in Volume 1, Part IV, present the criteria and many of the techniques that psychological researchers use to assess whether a measure is a good fit for a construct.

The Value of Multiple Operations

As you think about measures for a study, it is important to keep in mind that it is generally a good idea to include more than one operationalization of the constructs that interest you. Webb et al. (1999) set out the classic arguments for the value of having multiple operations to define the same underlying construct. They defined *multiple operationism* as the use of many measures that share a conceptual definition "but have different patterns of irrelevant components" (p. 35). Having multiple operations of a construct has positive consequences because

> once a proposition has been confirmed by two or more independent measurement processes, the uncertainty of its interpretation is greatly reduced. . . . If a proposition can survive the onslaught of a series of imperfect measures, with all their irrelevant error, confidence should be placed in it. (Webb et al., 1999, p. 35)

Of course, Webb and colleagues (1999) were quick to point out that our confidence in a finding is first and foremost accomplished by "minimizing error in each instrument and by a reasonable belief in the different and divergent effects of the sources of error" (p. 35) across the measures that we include.

An example shows how this works. Suppose in a study you measure the level of aggression between two people in three different ways: by unobtrusively observing participants' physical contact, by self-reported desire to harm one another, and by taking a physiological measure of arousal (one from Volume 1, Part III, Section 1; one from Section 2; and one from Section 5). You can be confident that these measures do not share irrelevant sources of error. Observed behaviors might be open to misclassification (a slap on the back might be coded as an act of aggression but really be one of friendship), but self-reports and physiological arousal less so. Self-reports are more open to responding in a socially desirable manner than unobtrusive observations or physiological measures. People become aroused by both love and hate but rarely self-report hate when love is the answer.

Now suppose your study was meant to test the hypothesis that the likelihood of aggression is related to state anxiety. If all three of your measures revealed similar predicted relations to measures of state anxiety or responded similarly to manipulations meant to increase state anxiety (say, the sounding of an alarm),[7] this would allow you to

[7]Of course, multiple measures of state anxiety are as desirable a feature of your study as multiple measures of aggression.

Copyright American Psychological Association

rule out the irrelevant influences (misinterpretation by observers, social desirability, etc.) on your three aggression measures as the cause of the relation. If results are inconsistent across operations, having the three measures allows you to speculate on what the important differences between operations might be and refine your theory about how the two constructs are related.

There is another benefit to including multiple operations in research. Predictions are often made from theories on the basis of presumed causal processes that include multiple steps, or causal linkages. These are the focus of causal modeling studies (Volume 2, Part IV), and they also pertain to other research designs. So, we might speculate that hearing an alarm increases state anxiety by increasing uncertainty and physiological arousal. In turn, uncertainty and arousal increase the likelihood of aggression. By including a measure of arousal in your study, along with observed and self-reported aggression measures, your study also tests this mediating mechanism.

Of necessity, the chapters in Volume 1, Part III, present their array of measurement methods largely as discrete choices. And, in Volume 1, Part IV, the methods for appraising a measure's fit with the construct of interest largely address the adequacy of each measure separately. But as you design a study, you should not think that you must choose one measure or another. Instead, when you consider which measure best captures your construct of interest, remember that no measure is perfect. And, to the extent reasonable, the more measures—that do not share the same imperfections and that test more than one linkage if a causal chain is hypothesized—the better.

QUANTITATIVE DATA ANALYSIS

Once you have chosen your design and measures, the next question you must answer is, "What data analysis procedure corresponds to your research design and the characteristics of your measures?" Volume 3, Part I, presents a compendium of your options for analysis. Section 1 presents different techniques that you can use to get to know your data as a whole. Special attention is paid to discovering distributional or other characteristics of data (e.g., outliers, missing values) that might dictate your approach to analysis or that might need to be addressed before your more substantive analysis should proceed. Section 2 presents some approaches to describing your data and techniques that you might use to communicate your findings to others.

Volume 3, Part I, Sections 3 to 5, presents both basic and advanced techniques for analyzing and interpreting social and behavioral science data. The chapters are organized first according to the number of dependent or outcome measures in the analysis. In Section 3 you will find methods that are used when you want to relate a single dependent or outcome variable (e.g., self-reported aggression) to one or more independent or predictor variables (e.g., state anxiety, age, sex).

In Volume 3, Part I, Section 4, statistical techniques are presented that apply to studies involving, in their most basic case, a single outcome or criterion measure that has been measured more than once, over time. So, if the chapter on "Collecting Longitudinal Data: Present Issues and Future Challenges" (Volume 2, Part III, Chapter 18) is of interest to you, you will likely find an analytic strategy here that meets your needs.

The methods in Volume 3, Part I, Section 5, pertain to data analyses in which you have many measures and make no distinction between which variables are independent or dependent, predictor or criterion. Most frequently, these techniques are used to

Copyright American Psychological Association

uncover the abstract, or latent, variables that underlie a set of observed, or manifest, variables. For example, we might use an exploratory factor analysis to determine the number of factors underlying a multi-item measure of aggression or a confirmatory factor analysis (on the same instrument) to test the theory that the tendency toward physical and verbal aggression are independent (i.e., knowing how likely people are to hit you tells you nothing about how likely they are to call you a nasty name). Within Volume 3, Part I, Sections 3 to 5, you will also find a distinction in analytic choices depending on whether they pertain to (a) variables that are measured continuously or (b) variables that place people (or other units) into classes. Some of these techniques take into account the categorical nature of variables in your analyses, whereas others help you to discover what these categorical distinctions among respondents might be.

So, to find the chapters of greatest interest to you in these sections, first you will need to answer three questions: (a) Does your research problem distinguish independent or predictor variables from dependent or outcome variables? (b) Are your variables measured over time? (c) Are your variables continuous or categorical, either in how they are conceptualized or how they are measured? Of course, in many instances your answers to these questions will be complex. Within the same study, for example, you might want to reduce the number of dependent variables by constructing a composite (e.g., factor analyze the observation, self-report, and physiological measures of aggression) and then use the composite in a regression analysis (one dependent variable) with other, multiple indicators. Or, you will have some continuous and some class variables. As you read the chapters, you will find that one great advance in statistical science has been the recent development of sophisticated techniques that permit the integrated analysis of data that vary in their characteristics and the number of variables involved.

Volume 3, Part 1, Section 6, presents methods for studies that take into account (and capitalize on) the interdependence of responses from multiple participants who are in interaction with one another. The units of analysis can be anywhere from a dyad (e.g., a husband and wife) to a large group (e.g., a sports team, workers in an office). Because the responses of individuals in these networks are dependent on the responses of others, the data analysis strategy must take this into account. Equally important, sometimes the research question focuses on the nature of the interdependence.

Finally, in Volume 3, Part I, Section 7, you will find three chapters that present some of the special issues that arise, and some special statistical techniques that are used, when researchers reanalyze or integrate data that was collected by others. In one instance, secondary data analysis, you work with raw data that might have been collected for another purpose (e.g., arrest rates and climate data in cities are used to test the relation between crime and temperature). In the other instance, meta-analysis, the statistical results of previous research becomes the raw data in a quantitative research synthesis.

Effect Sizes, or Relation Strength, and Their Interpretation

The chapters in Volume 3, Part I, look remarkably different from those that would have appeared *in* such a work a generation ago, far more so than the other parts of the handbook. Before the past three decades, testing the null hypothesis was the gold standard for drawing inferences about whether data revealed significant relations. More recently, the exclusive use of null hypothesis significance testing has become controversial, with some arguing that the practice should be abandoned entirely (Cohen, 1994). An American Psychological Association (APA) task force recommended that researchers need to ask

Copyright American Psychological Association

of their data not only, "Are these variables related, yes or no?" but also "How strong of a relationship is there?" (Wilkinson & Task Force on Statistical Inference, 1999).

Prominent among the methods used to describe data is the estimation and interpretation of *effect sizes*, or "the degree to which the phenomenon is present in the population" (Cohen, 1988, p. 9). With the coming of age of effect size estimation came the importance of understanding the difference between statistical significance and *clinical* or *practical* significance. The latter requires extrastatistical interpretation of the data. To assess practical significance, researchers (and others) must wrestle with the question of how strong a relation needs to be before it can be deemed meaningful or important (see Cooper, 2008).

The answer to this question always depends on the context in which the research takes place. Cohen (1988) suggested some general definitions for small, medium, and large effect sizes in the social sciences. In defining these adjectives, he compared different average effect sizes that he had encountered across disciplines in the behavioral sciences. However, Cohen did not intend his labels to serve as guides for the *substantive* interpretation of relations by social scientists. Rather, he intended his rules to assist with power analyses in planning future studies, a very different objective. Using Cohen's definitions to interpret the substantive importance of an effect size misapplies his work.

In fact, there is no fixed scale for the substantive interpretation of the size of a relation, and there is no substitute for knowing the research context of the specific question. Here is a simple example. Assume that we have the results of a study that evaluated an intervention that was conducted with 200 participants, 100 each in the intervention and control condition, and a dichotomous measure of success or failure. Using Cohen's definitions, an increase in success rate from 45% in the control condition to 55% in the intervention condition would be considered a small effect (equivalent to $r = .10$ explaining 1% of the variance). However, what if this effect were found on a measure of "suicides among previous attempters" and the intervention was access to an online psychological services? Personally, I would not be inclined to label this effect "small," practically speaking. However, if the study measured whether previous suicide attempters did or did not endorse the statement "life is worth living" after a year of daily psychotherapy "small effect" certainly would come to my mind.

In this example, I tried to demonstrate that the interpretation of effect sizes rests heavily on (a) the intrinsic value placed on the outcome variable (how valuable is even a small difference?) and (b) the cost of the intervention. Also, when interpreting the magnitude of effects, it is informative to use contrasting elements that are closely related to the topic at hand. For example, what other interventions have been used to prevent suicide among previous attempters? If the hotline and several other interventions have been tried and found to have no effect, suddenly the daily therapy effect starts to look larger, worth pursuing further.

Effect sizes also need to be interpreted in relation to the methodology used in the primary research. So, studies with more intensive treatments (e.g., more frequent therapy sessions), more sensitive research designs (within-subject rather than between-subject), and measures with less random error can be expected to reveal larger effect sizes, all else being equal.

Although null hypothesis testing is not ignored, the contents of this handbook clearly demonstrate the shifting of emphasis from "yes or no?" to "how much?" questions. In all of the chapters on data analysis, you will find a primary emphasis on estimating and interpreting the magnitude of relations.

Copyright American Psychological Association

Promoting Transparency in Research Methods

Over the past decade, psychological scientists have increasing come to embrace the tenets of open science, a movement meant to "increase openness, integrity and reproducibility of research" (Center for Open Science, 2020). Among the changes in practice espoused by the open science movement are (a) the registration of study methods and analysis strategies before a study has begun collecting data, (b) the sharing of data once it has collected, (c) complete and transparent reporting of research methods and data, and (d) direct access for the public and policy makers to research articles and results, regardless of their economic circumstance. Open science practices are meant to facilitate the replication of study methods and results and to allow for more complete evaluations of the strength and weaknesses of research. With greater transparency and accessibility will come greater trust in and increased use of the findings of psychological science.

The question you must answer about reporting your research is: "What do readers need to know about your study so they can (a) evaluate its trustworthiness; (b) replicate it, if they wish; and (c) use it along with other studies to synthesize research on your topic?" Note that this question uses the word *need* rather than the word *want*. This is because the standards for reporting research have become considerably more detailed in the past decade. This is especially true since the publication of the seventh edition of the APA's *Publication Manual* (APA, 2020). The *Publication Manual* includes tables and figures containing the journal article reporting standards (or the JARS; Appelbaum et al., 2018) that summarize the information editors, reviewers, and readers expect to see in your work.

Why are reporting standards needed? Two developments in psychology—indeed, in all the behavioral, social, and medical sciences—have led to an increased emphasis on complete research reporting. First, social science evidence is being increasingly used in public policy decision making. This use places new importance on understanding how research was conducted and what it found (APA Presidential Task Force on Evidence-Based Practice, 2006). Policy makers and practitioners who wish to make decisions that are informed by scientific evidence want to know how reliable the information they are considering is and in what context the data were collected. This dictates that research reports be comprehensive in their descriptions of methods and results.

Second, psychological scientists studying basic processes have found that as evidence about specific hypotheses and theories accumulates, greater reliance is being placed on syntheses of research, especially meta-analyses (see Valentine et al., Volume 3, Chapter 24, this handbook). Psychologists who use meta-analysis summarize findings, but they also use variations in research methods to find clues to the variables that might mediate differences in findings. These clues emerge by grouping studies on the basis of distinctions in their methods and then comparing their results. For example, a meta-analyst might group studies of the relation between state anxiety and aggression depending on what type of aggression measure was used: observation, self-report, or physiological. What are the implications if only one type of measure reveals a relation? This synthesis-based evidence is then used to guide the next generation of problems and hypotheses to be studied in new data collections. Meta-analysis requires detailed descriptions of what you have done. Without complete reporting of methods and results, the utility of your study is diminished.

In Cooper (2018), I introduced each item listed in the JARS as well as an explanation of why it was deemed important for inclusion. The bottom line is that without complete reporting, the value of your study for the users of your findings will be diminished.

Copyright American Psychological Association

The final part of the handbook addresses issues related to open science. It includes three chapters that address data management and the integrity of the research reporting process. The chapters in Volume 3, Part II, look at how to plan for managing your data, including documentation and storage so that it is easily understood by others (Chapter 25), what practices in data analysis that should be avoided so that your results are more easily replicable (Chapter 26), and a broader overview of ethical issues that arise in research not associated with the treatment of research participants (Chapter 27). All these chapters are new to the second edition of the handbook.

CONCLUSION

"Introductions" to edited works are required to address what the chapters that follow have in common as well as how the differences between chapters are to be understood. Writing such a chapter for a handbook as broad in scope as this one has required that I touch on similarities at lofty levels of abstraction, such as the methods by which people know things and the relation of science to democracy.

But I have been able as well to uncover some very down-to-earth examples of similarities in the chapters. For example, as I have noted several times, a principle shared by all of the authors is that the research methods you choose should be appropriate to answer the question that you pose. This dictum seems almost too obvious to state. Let us not fool ourselves, however. The opposing desire to use the tool you know even if it's not a perfect fit is often hard to resist. Hopefully, this handbook will expand your toolbox so that this latter approach loses its appeal.

Describing the differences between chapters and how they can be understood has presented an equally formidable challenge. It was easy to begin with the sequence of method choices—assessing the ethics and feasibility of different approaches, then choosing measures, a research design, statistical techniques, and ending with research reports—although we know that in practice these choices are never as linear as they appear in books.

Bringing an equally linear order to the array of research designs, measurements, and analytic techniques available to psychological scientists was the most difficult task. Different approaches to psychological research begin with different epistemic assumptions and then travel through subdisciplines with different traditions. Like the species that we study, the methods used by psychological scientists defy simple categorization. But this is a good thing (even if it causes trouble for editors). After all, if science is humankind's greatest achievement (and I think it is), then isn't turning the lens of science on ourselves the ultimate expression of our uniqueness?

Harris Cooper
Editor-in-Chief

References

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.).

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, *61*(4), 271–285. https://doi.org/10.1037/0003-066X.61.4.271

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and

Copyright American Psychological Association

Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Center for Open Science. (2020). COS: Center for Open Science. https://www.cos.io/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Cooper, H. (2007). *Evaluating and interpreting research syntheses in adult learning and literacy*. National Center for the Study of Adult Learning and Literacy. https://doi.org/10.1037/e549792010-001

Cooper, H. (2008). The search for meaningful ways to express the effects of interventions. *Child Development Perspectives*, *2*(3), 181–186. https://doi.org/10.1111/j.1750-8606.2008.00063.x

Cooper, H. (2018). *Reporting quantitative research in psychology: How to meet journal article reporting standards*. American Psychological Association. https://doi.org/10.1037/0000103-000

Feibleman, J. K. (1969). *An introduction to the philosophy of Charles S. Peirce*. MIT Press.

Ferris, T. (2010). *The science of liberty*. HarperCollins.

Godfrey-Smith, P. (2021). *Theory and reality: An introduction to the philosophy of science* (2nd ed.). University of Chicago Press. https://doi.org/10.7208/chicago/9780226771137.001.0001

Hume, D. (1978). A treatise on human nature. Oxford University Press. (Original work published 1739–1740)

Kounios, J., & Beeman, M. (2009). The *Aha!* moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science*, *18*(4), 210–216. https://doi.org/10.1111/j.1467-8721.2009.01638.x

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). University of Chicago Press. https://doi.org/10.7208/chicago/9780226458106.001.0001

Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher*, *33*(2), 3–11. https://doi.org/10.3102/0013189X033002003

Merton, R. K. (1957). Priorities of scientific discovery. In N. Storer (Ed.), *The sociology of science: Theoretical and empirical investigations* (pp. 635–659). University of Chicago Press.

Merton, R. K. (1979). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.

Murray, S., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, *9*, 429–434. https://doi.org/10.1038/nn1641

Neimark, E. D., & Estes, W. K. (1967). *Stimulus sampling theory*. Holden-Day.

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, *65*(1), 1–12. https://doi.org/10.1037/a0018326

Rutjens, B. T., Heine, S. J., Sutton, R. M., & van Harrevld, F. (2018). Attitudes toward science. *Advances in Experimental Social Psychology*, *57*, 125–165. https://doi.org/10.1016/bs.aesp.2017.08.001

Shweder, R. A. (1996). Quanta and qualia: What is the "object" of ethnographic research? In R. Jessor, A. Colby, & R. A. Shweder (Eds.), *Ethnography and human development: Context and meaning is social inquiry* (pp. 175–182). University of Chicago Press.

Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1999). *Unobtrusive measures*. SAGE.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(8), 594–604. https://doi.org/10.1037/0003-066X.54.8.594