

An Applied Guide to Research Designs

Quantitative, Qualitative, and Mixed Methods

Second Edition

W. Alex Edmonds
Thomas D. Kennedy

Nova Southeastern University



Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne



FOR INFORMATION:

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

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

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

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

Acquisitions Editor: Leah Fargotstein
Editorial Assistant: Yvonne McDuffee
eLearning Editor: Katie Anчета and
Robert Higgins
Production Editor: Olivia Weber-Stenis
Copy Editor: Amy Harris
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Jeff Bryant
Indexer: William Ragsdale
Cover Designer: Candice Harman
Marketing Manager: Susannah Goldes

Copyright © 2017 by SAGE Publications, Inc.

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

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Names: Edmonds, W. Alex, author. | Kennedy, Thomas D., author.

Title: An applied guide to research designs : quantitative, qualitative, and mixed methods / W. Alex Edmonds, Nova Southeastern University, Thomas D. Kennedy, Nova Southeastern University.

Other titles: Applied reference guide to research designs

Description: Los Angeles : SAGE, [2017] | Earlier edition published in 2013 as: An applied reference guide to research designs : quantitative, qualitative, and mixed methods. | Includes bibliographical references and index.

Identifiers: LCCN 2015045991 | ISBN 9781483317274 (pbk. : alk. paper)

Subjects: LCSH: Social sciences—Research—Methodology. | Education—Research—Methodology. | Research—Methodology.

Classification: LCC H62 .E327 2017 | DDC 001.4/2—dc23
LC record available at <http://lccn.loc.gov/2015045991>

This book is printed on acid-free paper.

16 17 18 19 20 10 9 8 7 6 5 4 3 2 1

Brief Contents

List of Tables, Figures, and Diagrams	xi
Foreword	xiv
Preface	xvii
Acknowledgments	xxv
About the Authors	xxvi
1. A Primer of the Scientific Method and Relevant Components	1
PART I. QUANTITATIVE METHODS FOR EXPERIMENTAL AND QUASI-EXPERIMENTAL RESEARCH	29
2. Between-Subjects Approach	35
3. Regression-Discontinuity Approach	51
4. Within-Subjects Approach	57
5. Factorial Designs	75
6. Solomon <i>N</i> -Group Designs	93
7. Single-Case Approach	103
PART II. QUANTITATIVE METHODS FOR NONEXPERIMENTAL RESEARCH	117
8. Ex Post Facto and Posttest-Only Designs	120
9. Observational Approach	125
10. Survey Approach	133

PART III. QUALITATIVE METHODS	141
11. Grounded Theory Perspective	145
12. Ethnographic Perspective	152
13. Narrative Perspective	160
14. Phenomenological Perspective	168
PART IV. MIXED METHODS	177
15. Convergent-Parallel Approach	181
16. Embedded Approach	189
17. Explanatory-Sequential Approach	196
18. Exploratory-Sequential Approach	201
19. Mixed Methods, Case Studies, and Single-Case Approaches	208
20. Action Research Approaches	212
21. Conclusion	218
Appendix A. Less Common Designs for Experimental and Quasi-Experimental Research	222
Appendix B. Types of Case Study Designs	229
Appendix C. Five Types of Mixed Methods Designs	231
Appendix D. Reporting Preliminary Statistical Findings	232
Appendix E. Analysis by Design	251
Appendix F. Qualitative Data Analysis	321
Appendix G. Validity, Control, and Causal Inferences Checklist	332
Appendix H. Consort 2010: Checklist and Guidelines for Reporting Parallel Group Randomised Trials	334
References	338
Author Index	352
Subject Index	357

Detailed Contents

List of Tables, Figures, and Diagrams	xi
Foreword	xiv
<i>Paul E. Spector</i>	
Preface	xvii
Audience	xx
Features	xxi
Unique Features	xxi
New to the Second Edition	xxii
Companion Website	xxiii
Acknowledgments	xxv
About the Authors	xxvi
1. A Primer of the Scientific Method and Relevant Components	1
The Scientific Method	2
Validity and Research Designs	3
<i>Independent and Dependent Variables</i>	5
<i>Internal Validity</i>	7
<i>External Validity</i>	8
<i>Construct Validity</i>	9
<i>Statistical Conclusion Validity</i>	9
Design Logic	13
<i>Control</i>	13
Design Notations	15
Assignment Techniques	15
Comparison and Control Groups	18
Sampling Strategies	18
<i>Probability Sampling Techniques</i>	20
<i>Nonprobability Sampling Techniques</i>	20

Research Questions	21
<i>Types of Inquiry</i>	21
PART I. QUANTITATIVE METHODS FOR EXPERIMENTAL AND QUASI-EXPERIMENTAL RESEARCH	29
Experimental Research	30
Quasi-Experimental Research	32
2. Between-Subjects Approach	35
Pretest and Posttest Designs	35
<i>k-Factor Designs</i>	36
Posttest Designs	43
3. Regression-Discontinuity Approach	51
4. Within-Subjects Approach	57
Repeated-Measures Approach	58
<i>Switching-Replications Design: A Primer</i>	59
<i>Crossover and Latin-Square Designs</i>	61
5. Factorial Designs	75
6. Solomon <i>N</i> -Group Design	93
7. Single-Case Approach	103
PART II. QUANTITATIVE METHODS FOR NONEXPERIMENTAL RESEARCH	117
8. Ex Post Facto and Posttest-Only Designs	120
9. Observational Approach	125
Explanatory Design	126
Predictive Design	126
<i>Statistical Tools</i>	127
<i>Predicting a Single Variable for the Predictive Design</i>	128
<i>Predicting More Than One Variable</i>	128
10. Survey Approach	133
Primary Threats to Validity	134
Cross-Sectional Design	135
Longitudinal Design	135

PART III. QUALITATIVE METHODS	141
The Case Study	143
11. Grounded Theory Perspective	145
Systematic Design	146
Emerging Design	146
Constructivist Design	146
12. Ethnographic Perspective	152
Realist Design	152
Critical Design	153
Case Study Design	153
13. Narrative Perspective	160
Descriptive Design	161
Explanatory Design	161
Critical Design	161
14. Phenomenological Perspective	168
Existential Design	168
Transcendental Design	169
Hermeneutic Design	169
Case Study Design	169
PART IV. MIXED METHODS	177
Mixed Methods Legend	179
15. Convergent-Parallel Approach	181
Parallel-Databases Design	182
Data-Transformation Design	182
Data-Validation Design	182
Multilevel Design	182
16. Embedded Approach	189
Embedded-Experiment Design	190
Embedded-Correlational Design	190
Embedded Case Study Design	190
17. Explanatory-Sequential Approach	196
Follow-Up Explanations Design	197
Participant-Selection Design	197

18. Exploratory-Sequential Approach	201
Instrument-Development Design	202
Theory-Development Design	202
Treatment-Development Design	202
19. Mixed Methods, Case Studies, and Single-Case Approaches	208
Mixed Method A-B-A Designs	209
<i>Sequential Case Study Single-Case Design</i>	210
20. Action Research Approaches	212
Embedding Research Designs Within the Action Research Approach	214
21. Conclusion	218
Philosophical Tenets	218
Evaluation Approaches	219
Data Analytics	220
Final Remarks	220
Appendix A. Less Common Designs for Experimental and Quasi-Experimental Research	222
Appendix B. Types of Case Study Designs	229
Appendix C. Five Types of Mixed Methods Designs	231
Appendix D. Reporting Preliminary Statistical Findings	232
Appendix E. Analysis by Design	251
Appendix F. Qualitative Data Analysis	321
Appendix G. Validity, Control, and Causal Inferences Checklist	332
Appendix H. Consort 2010: Checklist and Guidelines for Reporting Parallel Group Randomised Trials	334
References	338
Author Index	352
Subject Index	357

List of Tables, Figures, and Diagrams

TABLES ◆

- | | |
|---|---|
| Table 1.1. Threats to Internal Validity 7 | Table 1.6. Design Notations 15 |
| Table 1.2. Threats to External Validity 9 | Table 1.7. Group Assignment Design Notations 15 |
| Table 1.3. Threats to Construct Validity 10 | Table 7.1. Design Notations for Single-Case Approaches 104 |
| Table 1.4. Threats to Statistical Conclusion Validity 11 | Table 20.1. Design Notations for the Action Research Approach 215 |
| Table 1.5. Control Techniques for Experimental, Quasi-Experimental, and Nonexperimental Research 14 | |

FIGURES ◆

- | | |
|--|---|
| Figure 1.1. The Scientific Method 3 | Figure 3.1. Sample of a Cutoff Score 52 |
| Figure 1.2. Conceptual and Operational Definitions 6 | Figure 4.1. Wait-List Continuation Design (for Random Assignment) 60 |
| Figure 1.3. Example of a Stratified Random-Assignment Technique 17 | Figure 4.2. Switching-Replications Design (for Nonrandom Assignment) 60 |
| Figure 1.4. Example of a Sample Extracted From a Population 19 | Figure 11.1. Systematic Design 147 |
| Figure I.1. Quantitative Method Flowchart 29 | Figure 11.2. Emerging Design 148 |
| Figure I.2. Double Pretest Design for Quasi-Experimental Research 33 | Figure 11.3. Constructivist Design 150 |
| | Figure 12.1. Realist Design 155 |
| | Figure 12.2. Critical Design 156 |

- Figure 12.3. Case Study Design 157
- Figure 13.1. Descriptive Design 163
- Figure 13.2. Explanatory Design 164
- Figure 13.3. Critical Design 165
- Figure 14.1. Existential Design 171
- Figure 14.2. Transcendental Design 172
- Figure 14.3. Hermeneutic Design 173
- Figure 14.4. Case Study Design 174
- Figure 15.1. Parallel-Databases Design 183
- Figure 15.2. Data-Transformation Design 184
- Figure 15.3. Data-Validation Design 185
- Figure 15.4. Multilevel Design 187
- Figure 16.1. Embedded-Experiment Design (One-Phase) 191
- Figure 16.2. Embedded-Correlational Design 192
- Figure 16.3. Embedded Case Study Design 193
- Figure 17.1. Follow-Up Explanation Design 197
- Figure 17.2. Participant-Selection Design 198
- Figure 18.1. Instrument-Development Design 203
- Figure 18.2. Theory-Development Design 204
- Figure 18.3. Treatment-Development Design 205
- Figure 19.1. Sequential Case Study Single-Case Design 210
- Figure 20.1. The Cyclical Action Research Approach 214

♦ DIAGRAMS

- Diagram 2.1. Pretest and Posttest Control Group Design 37
- Diagram 2.2. 2-Factor Pretest and Posttest Control Group Design 38
- Diagram 2.3. 2-Factor Pretest and Posttest Design 40
- Diagram 2.4. 3-Factor Pretest and Posttest Design 41
- Diagram 2.5. 4-Factor Pretest and Posttest Design 42
- Diagram 2.6. Example of a Posttest-Only and a Historical Control Group Design 44
- Diagram 2.7. Posttest Control Group Design 45
- Diagram 2.8. 2-Factor Posttest Control Group Design 46
- Diagram 2.9. 2-Factor Posttest Design 47
- Diagram 2.10. 3-Factor Posttest Design 48
- Diagram 3.1. Regression-Discontinuity Pretest-Posttest Control Group Design₁ 53
- Diagram 3.2. Regression-Discontinuity Pretest-Posttest Control Group Design₂ 54
- Diagram 4.1. A Repeated-Measures Approach 3-Factor Crossover Design 62
- Diagram 4.2. Example of a 3 × 3 Latin-Square Design 63

- Diagram 4.3. Pretest and Posttest Design (One-Group) 64
- Diagram 4.4. Pretest and Multiple-Posttest Design 65
- Diagram 4.5. Switching-Replications Design (Experimental) 66
- Diagram 4.6. Switching-Replications Design (Quasi-Experimental) 67
- Diagram 4.7. Crossover Design (2-Factor) 68
- Diagram 4.8. Interrupted Time-Series Design (One-Group) 69
- Diagram 4.9. Interrupted Time-Series Design (Two-Group) 70
- Diagram 4.10. 6×6 Latin-Square Design 72
- Diagram 5.1. 2×2 Factorial Pretest and Posttest Design (Between-Subjects) 77
- Diagram 5.2. 2×2 Factorial Posttest Design (Between-Subjects) 79
- Diagram 5.3. 2×2 Factorial Pretest and Posttest Design (Within-Subjects) 81
- Diagram 5.4. 3×2 Factorial Posttest Design (Between-Subjects) 82
- Diagram 5.5. $2 \times 2 \times 2$ Factorial Posttest Design (Between-Subjects) 85
- Diagram 5.6. $2 \times 3 \times 2$ Factorial Posttest Design (Between-Subjects) 87
- Diagram 5.7. $2 \times 2 \times 2$ Mixed-Factorial Design (Mixed-Subjects) 89
- Diagram 6.1. Solomon Four-Group Design 94
- Diagram 6.2. Solomon Six-Group Design 96
- Diagram 6.3. Solomon Eight-Group Design 99
- Diagram 7.1. A-B-A Design 106
- Diagram 7.2. A-B-A-B Design 107
- Diagram 7.3. A-B-A-B-C-B-C Design 108
- Diagram 7.4. Changing Criterion Design (A-B) 110
- Diagram 7.5. Multiple Baseline Design (A-B) 111
- Diagram 7.6. Multiple Baseline Design (A-B-C) 112
- Diagram 8.1. Two-Group Ex Post Facto Design 121
- Diagram 8.2. Posttest Design (One-Group) 121
- Diagram 8.3. Ex Post Facto Design 122
- Diagram 9.1. Explanatory Design 129
- Diagram 9.2. Predictive Design₁ 130
- Diagram 9.3. Predictive Design₂ 131
- Diagram 10.1. Cross-Sectional Design 136
- Diagram 10.2. Longitudinal Design 136
- Diagram 19.1. A-B-A Concurrent Design 209
- Diagram 19.2. A-B-A Sequential Design 210
- Diagram 19.3. A-B-A Nested Design 210
- Diagram 20.1. Posttest With a Historical Control for the Action Research Approach 215
- Diagram 20.2. Pretest and Posttest Design for the Action Research Approach 216
- Diagram 20.3. Posttest-Only Design for the Action Research Approach 216
- Diagram 20.4. Ex Post Facto Design for the Action Research Approach 216

Foreword

Research methods in the social sciences (I include behavioral sciences and education under this broader term) are quite varied, as the number of settings and situations in which our investigations occur are immense. Furthermore, concerns about research ethics and potential harm inflicted on human subjects put severe limits on what is feasible, requiring social scientists to have at their disposal a plethora of possible approaches to apply to a given problem. Social scientists, particularly those who do nonexperimental research and those who conduct field research in naturalistic settings, must be well versed in many varied methods in order to be prepared to apply optimal approaches to address their research questions in specific settings. Although different disciplines within social science have developed some of their own approaches and practices, there is a great deal of overlap among disciplines in the methodologies utilized.

One can classify methods according to whether they are qualitative or quantitative. Although there are a variety of approaches, qualitative methods are distinguished by their collection and synthesis of information in a largely nonquantitative way. For example, individuals might be interviewed about their experiences living in a college dormitory. The researcher will review and synthesize the responses, looking for meaningful themes. Quantitative methods, on the other hand, tend to involve defining variables in advance and then quantifying observations of those variables. Thus, one might ask college students to complete a survey in which questions ask for ratings about various aspects of the dormitory experience, such as how much they liked the food on a 1-to-5 scale. Different disciplines within social science tend to favor one approach over the other—for example, anthropologists make frequent use of qualitative methods whereas psychologists mainly use quantitative methods, although there are exceptions on both sides.

The quantitative methods training in social science often divides methodology into assessment, design, and statistics. Assessment concerns

techniques for measuring or operationalizing variables largely using multiple-item psychological tests and scales. Design is the structure of an investigation that defines the sequence and nature of both the conditions subjects are exposed to and the observations taken on those subjects. Statistics concerns the mathematical procedures used to analyze the quantitative data produced by the study once the design is implemented. Assessment, design, and statistics are all important elements in any quantitative investigation, so students must develop expertise in all three. Certainly one must have sound measurement to be able to draw conclusions about the underlying variables of interest, and one must analyze data using appropriate statistics, but it is the *design* of the investigation that is most important in being able to draw inferences from an investigation.

This book deals primarily with design, including both designs for experimental and nonexperimental research. It is perhaps unique in providing a balanced treatment of both qualitative and quantitative methods that are integrated at the end when mixed methods are discussed. The book begins with a general discussion of basic principles of the scientific method in social science, including topics such as validity and control. It then covers quantitative methods, including experimental, quasi-experimental (in Part I), and nonexperimental (in Part II) research. Part III discusses a variety of qualitative methods, including grounded theory, ethnography, narrative, and phenomenological approaches. The book concludes (Part IV) with a treatment of mixed methods and action research that involves elements of both qualitative and quantitative methods.

This book includes both basic and advanced designs, making it useful as both a textbook for students in a course that covers design and as a guide to experienced researchers. The book provides an example from the literature for every design covered. A brief overview is provided of each example study's research question and procedure, as well as recommended statistical approaches for data analysis. The citation is provided from widely available journals, so each article can be consulted for more details. Thus, the reader can easily see how each research team was able to use each design and how those researchers handled data analysis and interpretation.

Although not every social scientist will use all of these designs, the serious student of social science research methodology needs a basic understanding of how design features inform appropriate inference. Such a student should have a working knowledge of qualitative and quantitative methods, as provided in this book. It can provide an introduction to design that can later serve as a reference to details of specific designs that can be applied to a particular problem. The development of the computer over the

past three decades has shifted much of the focus of social scientists, especially those who do quantitative studies, from design to statistics as increasing computing power has allowed the development of increasingly computationally complex statistical methods. Thus, in graduate programs we find many classes on statistics but few on design. It is important to remember that it is the *design* and *not* the statistic that is the basis for inference, making the study of design of vital importance, and this book is an invaluable resource for both social scientists and aspiring social scientists.

Paul E. Spector
University of South Florida

Preface

The objective of this reference book is to visually present, with consistent terminology, quantitative, qualitative, and mixed methods research designs in education and the social and behavioral sciences in a way that students and researchers can readily understand and accurately apply in their own investigations. Through our experience and research for this guide, we realized there are many inconsistencies and variations of terminology, both within and between research texts in education and the social and behavioral sciences, especially with the use of the terms *method*, *research*, *approach*, and *design*. We believe that the terminology should be clearly distinguished with the appropriate nomenclature. The interchange of terminology creates confusion among consumers of research, particularly students. We attempt to resolve the confusion by breaking down each aspect of the research terminology into its components in a hierarchical fashion to provide clarity for the reader. As seen in the chart that follows, the resulting nomenclature is thus used throughout the text for the quantitative and the qualitative method.

QUANTITATIVE	
<i>Level</i>	<i>Explanation</i>
METHOD ₁	The <i>method</i> provides the theoretical, philosophical, and data-analytic stance (e.g., a quantitative method ₁).
▼	▼
RESEARCH ₂	<i>Research</i> refers to the systematic process of control (e.g., group assignment, selection, and data collection techniques). Research can be experimental, quasi-experimental, or nonexperimental (e.g., a quantitative method ₁ and experimental research ₂).

(Continued)

(Continued)

QUANTITATIVE	
<i>Level</i>	<i>Explanation</i>
▼	▼
APPROACH ₃	The <i>approach</i> is the first step to creating structure to the design, and it details (a) a theoretical model of how the data will be collected and (b) if one case, one group, or multiple groups will be associated with the process (e.g., a quantitative method ₁ , experimental research ₂ with a between-subjects approach ₃).
▼	▼
DESIGN ₄	The <i>design</i> is the actual structure or framework that indicates (a) the time frame(s) in which data will be collected, (b) when the treatment will be implemented (or not), and (c) the exact number of groups that will be involved (e.g., a quantitative method ₁ , experimental research ₂ with a between-subjects approach ₃ and a pre- and posttest control group design ₄).

QUALITATIVE	
<i>Level</i>	<i>Explanation</i>
METHOD ₁	The <i>method</i> provides the theoretical, philosophical, and data analytic stance (e.g., a qualitative method ₁).
▼	▼
RESEARCH ₂	<i>Research</i> for the qualitative method is nonexperimental (e.g., a qualitative method ₁ and nonexperimental research ₂).
▼	▼
PERSPECTIVE ₃	The <i>perspective</i> is the first step to creating structure to the design, and it details the theoretical perspective (or lens) of how the researcher(s) will approach the study (e.g., a qualitative method ₁ , nonexperimental research ₂ with an ethnographic perspective ₃).
▼	▼
DESIGN ₄	The <i>design</i> is the actual structure that indicates (a) if one case, one group, or multiple groups will be associated with the process and (b) when the data will be analyzed (e.g., a qualitative method ₁ , nonexperimental research ₂ with an ethnographic ₃ and a case study design ₄).

This book is designed to improve one's ability to conceptualize, construct, test, problem solve, and acquire knowledge, all of which are characteristics of scientific inquiry and the creative process required when

conducting research. We have discovered, in teaching research methods courses and supervising dissertation committees, that students often have a difficult time conceptualizing the most appropriate research design, en route to collecting the data and answering the stated research questions or hypotheses. Based on this observation, we sought to find the best text that could help resolve this critical issue. We found that most research methods texts are more broad, covering the entire spectrum of the research process while devoting only a single chapter or a few sections embedded throughout (often incomplete) to research designs. Furthermore, the authors of these texts often present the research designs without quality visual representations and sound real-life examples. The issue is further confounded when investigators omit from the Methods portion of published manuscripts an accurate description of the research design that was employed.

We also discuss the issue of inconsistent terminology; for example, it is not unusual to see authors use the following terms interchangeably: a *correlational method*, *correlational research*, a *correlational approach*, or a *correlational design*. Although this may not be entirely wrong, it is not entirely accurate (or specific) and can lead to confusion. It would be more accurate to say all together in sequence, a quantitative method, nonexperimental research, an observational approach, and a predictive design (and then, of course, the correlational statistic or regression analysis is applied to the observational data). These inconsistencies, at best, can lead to confusion and difficulties when attempting to conceptualize and choose the design that best fits the research problem and subsequent questions or hypotheses; at worst, they can render findings invalid. Considering these aspects, students often find themselves lost at that critical part of the research process while attempting to (a) choose a design that will allow for the acquisition of data best suited to answer their research questions or examine their hypotheses and (b) incorporate a design into the Procedures section of their Methods chapter of their dissertation.

Although we cover a variety of practical research designs in quantitative, qualitative, and mixed methods, this book is not intended to be a complete reference guide for individuals conducting program evaluations. We briefly address this issue at the end of this guide. We revealed through our research for this book that many of these sound research designs presented within are underused in education and the social and behavioral sciences. We hope the presentation of these materials will continue to strengthen research in the area through the application of sound methodological techniques. These designs can be applied in field, laboratory, and even web-based settings. Although this book does not go into great detail regarding the theory or philosophy of qualitative or quantitative methods

and the associated research designs, we do provide recommended texts and articles for the reader who is interested in a more thorough understanding of a particular approach, method, or design. The intent of this book is seemingly paradoxical, in that we attempt to give students and researchers a dense (no filler) yet quick reference guide for conceptualizing and creating a design that best fits the primary research question. Thus, this is an applied text, using visual aids and real-world examples, rather than covering foundational and theoretical issues.

Visually delivering the information coupled with relevant examples may optimize the learning process and subsequent application of learning. The reader will notice that we often state that further decisions about the applications of particular research designs should be based on “theoretical and logistical considerations.” Although we attempt to apply linear logic and black-and-white elements to scientific methodology, there are many instances in which “rules of thumb” and old laws do not apply. Research in the field of social sciences is still relatively new, and the vast and varied contexts in which we investigate create a level of complexity and sophistication that often requires subjectivity and interpretation.

As mentioned earlier, this book is meant to cover the most practical and common research designs currently used in educational and the social and behavioral sciences. Referring to these research designs as “common” or “practical” is somewhat a misnomer, and it does not imply that the designs are less powerful or the results have less meaning. In reviewing many articles over the years, we have noticed that, all too often, researchers use unnecessarily complex research designs that complicate the application and subsequent statistical analyses, leaving much more room for error. *Parsimony* is a favorable word in science; that is, a design should be as complex as it needs to be and, at the same time, as simple as it needs to be.

♦ AUDIENCE

The target audience for this book is the researcher in the fields of education, sociology, psychology, nursing, and other human service fields. More specifically, this book is written for undergraduate students working on honors theses and for graduate students working on theses or dissertations. This book will assist all students who (a) have a basic understanding of research methods, (b) are in the process of conducting research, or (c) plan on conducting research at some point during their careers. Furthermore, it can also be used as a tool by professors who are either teaching research courses or supervising students on theses and dissertations. Specifically, professors will

find this reference guide useful in assisting and guiding students interested in improving their understanding of how research is set up and conducted. We have attempted to create a visual system in the form of a practical, easy-to-follow reference guide to help in the conceptualization and development of many of the common research designs. We offer examples of this visual system “in action” for each of the designs presented with the use of published studies. In addition, further recommendations and suggestions are provided for those interested in acquiring a more comprehensive understanding of basic research designs. The book includes the core designs that are used by quantitative, qualitative, and mixed method researchers.

FEATURES ♦

We have maintained a singular focus on research designs and have provided an example for each chosen research design with a relevant peer-reviewed article. We have incorporated a number of features throughout the book that will assist students in designing their own studies. In each example, we summarize the procedures and include the relevant variables in a design notation model. In many cases, the research article examples include multiple research questions or hypotheses. However, for the sake of clarity, we attempt to present one overarching research question that summarizes the major goal of the study. In some instances, for research in qualitative methods, we include a research aim. Research using mixed methods usually contains an additional research question to answer the inquiry associated with combining quantitative results and qualitative findings. In addition to the example research designs, we also include brief discussions on (a) the relevant aspects of research, (b) different types of designs, (c) the scientific method, and (d) a list of recommended readings pertaining to each area. Also, located in the appendices, we present many examples of rarely applied research designs, as well as case study designs, with brief notations on the intended use and effectiveness.

UNIQUE FEATURES ♦

We have summarized and condensed over 140 articles and books included in this reference guide. In addition, there are many unique features associated with this guide. These features were included to enhance the understanding of the concepts and designs presented:

- An array of relevant references and sources for the reader
- Consistent terminology, which is emphasized throughout (a standardized taxonomy)
- Discussions on the differences between within- and between-subject approaches
- The use of the k -factor design as a means to distinguish multiple-treatment groups
- Inclusion of both within-subjects and between-subjects k -factor designs
- Diagrams of factorial designs
- Examples of the Solomon N -group designs as an extension of the four-group design
- Diagrams of single-case approaches
- Diagrams of nonexperimental research such as observational (correlational) and survey approaches
- Visual models for qualitative methods
- Proposed designs for mixed method single-case approaches and the action research approach
- Appendices covering examples of rarely used, but relevant, research designs for experimental and quasi-experimental research, case studies, and mixed methods
- An appendix covering statistical analysis by design and a brief guide to reporting statistics
- An appendix devoted to coding qualitative data

♦ NEW TO THE SECOND EDITION

For the second edition of this book, we have expanded and added upon the already extensive coverage of research designs for qualitative, quantitative, and mixed methods. Many of the additions were included based on feedback from students, researchers, reviewers, and our own personal experiences using the text in classes and in our research endeavors. As a result, the coverage of research designs is more complete and includes topics not previously addressed. In addition, in many cases, we were able to update critical references in relation to a specific design or approach discussed within the text.

This is a general list of the topics and areas included in the second edition:

- Explanations and examples of conceptual and operational definitions for independent variables

- Examples of moving from the primary research question to the specific method, approach, and then appropriate research design
- Elaboration on the survey approach, including a discussion of the most common threats of external and construct validity
- Presentation of statistical tools common to the observational approach for nonexperimental research
- An expanded discussion on the switching-replication design, including a variant of this design when random assignment is used called the *wait-list continuation design*
- An expanded chapter and updated references for the single-case approach
- A new chapter on the action research approach, including diagrams for specific research designs that can be used for participatory action research approaches
- A new appendix that details specific statistical analysis in relation to the research design, which includes vignettes, research questions, the design, and step-by-step analysis using SPSS (data files available via the companion website)
- A new appendix covering the general procedural steps for qualitative data analysis
- A new appendix that presents quick reference materials for writing and summing up preliminary statistical analysis, including effect size estimations, power analysis, and statistical symbol interpretations
- A chart that allows researchers to gauge the quality of an experimental study based on internal validity, statistical conclusion validity, control, and cause-and-effect parameters
- The *Consort 2010: Checklist and Guidelines for Reporting Parallel Randomised Trials*

In relation to the updated content of the text, the ancillary materials were also updated. This includes the PowerPoints, syllabi, discussion questions, and assignments. Materials such as data sets, checklists, and articles are also available via the companion website.

COMPANION WEBSITE ♦

Visit the companion website at study.sagepub.com/edmonds2e to access valuable instructor and student resources. These resources include PowerPoint slides, discussion questions, class activities, SAGE journal articles, web resources, and online data sets.

Acknowledgments

We would like to thank Daniel L. Stufflebeam, PhD, and Ronald J. Chenail, PhD, for providing their insight, thoughts, and suggestions about our book. Dr. Stufflebeam and Dr. Chenail's expertise in their respective areas is undeniable and ultimately helped strengthen the overall product. We would also like to thank Paul E. Spector, PhD, for his insight and contribution to this book. His dedication to the field of scientific inquiry is unwavering, and we are honored to have him as a part of this project. We would like to thank Robert K. Yin, PhD, for his thoughts and suggestions, as well as Destine Pine for his editorial and formatting contributions. A special thanks goes to Vicki Knight at SAGE for her support and dedication to the project. Although Vicki retired following the production of the second edition, she provided a great deal of insight and thoughtful reviews that contributed greatly to this new edition. We wish her the best of luck. A number of reviewers provided useful and valuable feedback, which ultimately shaped the book: Jennifer E. Dalton, University of Toronto; Donna Schnorr, California State University, San Bernardino; Behzad Esmaeili, University of Nebraska–Lincoln; Rebekah Cole, Old Dominion University; David Han, University of Texas at San Antonio; Raphael Heaggans, Niagara University; Lydia Kyei-Blankson, Illinois State University; Michael R. Pryzdia, Arizona State University; Micheál Newell, National University of Ireland, Galway; Marnie L. Moist, Saint Francis University; Michael R. Pryzdia, Arizona State University; Fairuz J. Lutz, Wesley College; Janet Reid-Hector, Rutgers University; Angela M. Sikorski, Texas A&M University–Texarkana; Julia C. Tang, Mount Saint Mary's University, Los Angeles; Dennis L. Veleber, University of Great Falls.

About the Authors

W. Alex Edmonds, PhD, BCB, is an associate professor of research at Nova Southeastern University in Davie, Florida. He graduated from Florida State University and received his doctoral degree in Educational Psychology with a minor in Statistics and Measurement. Over the years, Dr. Edmonds has applied his knowledge of research design, measurement, and assessment in both field and laboratory examinations. He has published extensively in a variety of areas, such as psychophysiology and educational psychology. His primary interest revolves around applying unique methodological and statistical techniques as a means to exploring the relationship between emotions, physiology, and performance in a variety of domains. He also has over 10 years of experience in applying biofeedback in field and research applications. While in graduate school, he conducted his field work with the track and field team at Florida State and started using biofeedback for research and practice during this time. He has utilized biofeedback extensively with various types of athletes for performance enhancement, as well as stress-regulation techniques for individuals with type 2 diabetes and pain management for patients suffering from chronic pain. Dr. Edmonds is certified through the Biofeedback Certification International Alliance in general biofeedback.

Thomas D. Kennedy, PhD, BCB, is an associate professor of research at Nova Southeastern University in Davie, Florida. He received his PhD in Counseling Psychology from the University of Miami and his MA in Clinical Psychology from Southern Methodist University. For 5 years, he has been a faculty member at NSU, where he has developed and taught research and statistics courses. He is the Institutional Review Board representative for the School of Education. His clinical experience consists of providing neuropsychological assessments, behavioral medicine interventions, and group therapy in inpatient and outpatient settings, including the following: the University of Texas Southwestern Medical Center, University of Miami Mailman Center for Child Development, Jackson

Memorial Hospital, University of Miami Institute for Individual and Family Therapy, and the Dallas County Jail. His research interests include two convergent tracks, one focusing on at-risk children and adolescents and the other exploring biofeedback and other complementary and alternative medicine interventions. He has received multiple grants and provides evaluation and grant writing services for various organizations in southern Florida. He has published and presented in the areas of juvenile crime, psychophysiology, and research methodology. Dr. Kennedy is a licensed psychologist PY 8307 and is certified through the Biofeedback Certification International Alliance in general biofeedback; he maintains a small private practice for children and adolescents.

I'd like to dedicate this book to all my past supervisors, professors, teachers, and dedicated students. They have provided me the support, skills, interest, and motivation to continue the lines of scholarly work and achievement in a never-ending effort to expand our minds.

—Alex Edmonds

I would like to dedicate this book to my wife Karen and my three boys, Aiden, Ethan, and Ashton.

—Tom Kennedy

CHAPTER 1

A PRIMER OF THE SCIENTIFIC METHOD AND RELEVANT COMPONENTS

The primary objective of this book is to help researchers understand and select appropriate designs for their investigations within the field, lab, or virtual environment. Lacking a proper conceptualization of a research design makes it difficult to apply an appropriate design based on the research question(s) or stated hypotheses. Implementing a flawed or inappropriate design will unequivocally lead to spurious, meaningless, or invalid results. Again, the concept of validity cannot be emphasized enough when conducting research. Validity maintains many facets (e.g., statistical validity or validity pertaining to psychometric properties of instrumentation), operates on a continuum, and deserves equal attention at each level of the research process. Aspects of validity are discussed later in this chapter. Nonetheless, the research question, hypothesis, objective, or aim is the primary step for the selection of a research design.

The purpose of a research design is to provide a conceptual framework that will allow the researcher to answer specific research questions while using sound principles of scientific inquiry. The concept behind research designs is intuitively straightforward, but applying these designs in real-life

situations can be complex. More specifically, researchers face the challenge of (a) manipulating (or exploring) the social systems of interest, (b) using measurement tools (or data collection techniques) that maintain adequate levels of validity and reliability, and (c) controlling the interrelationship between multiple variables or indicating emerging themes that can lead to error in the form of confounding effects in the results. Therefore, utilizing and following the tenets of a sound research design is one of the most fundamental aspects of the scientific method. Put simply, the research design is the structure of investigation, conceived so as to obtain the “answer” to research questions or hypotheses.

♦ THE SCIENTIFIC METHOD

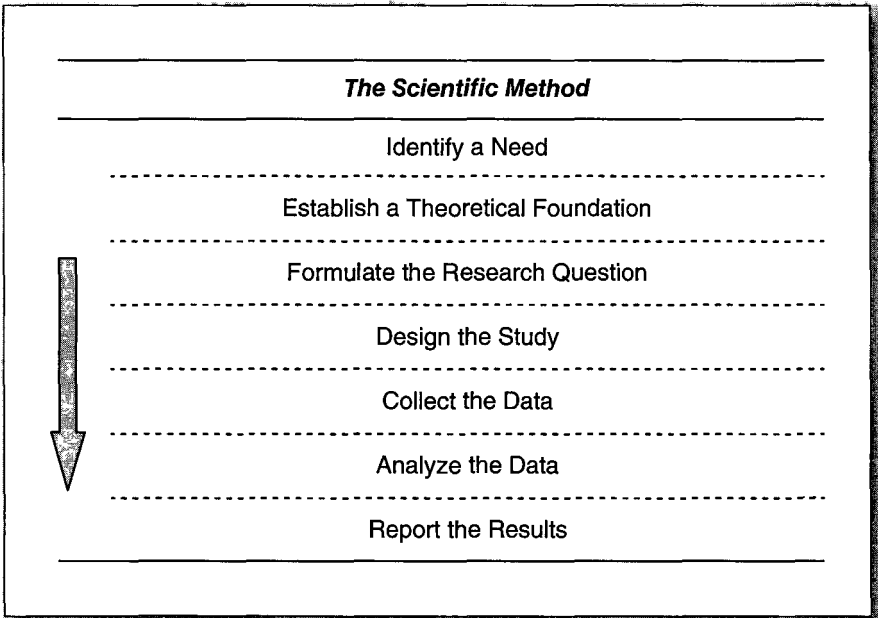
All researchers who attempt to formulate conclusions from a particular path of inquiry use aspects of the scientific method. The presentation of the scientific method and how it is interpreted can vary from field to field and method (qualitative) to method (quantitative), but the general premise is not altered. Although there are many ways or avenues to “knowing,” such as sources from authorities or basic common sense, the sound application of the scientific method allows researchers to reveal valid findings based on a series of systematic steps. Within the social sciences, the general steps include the following: (a) state the problem, (b) formulate the hypothesis, (c) design the experiment, (d) make observations, (e) interpret data, (f) draw conclusions, and (g) accept or reject the hypothesis. All research in quantitative methods, from experimental to nonexperimental, should employ the steps of the scientific method in an attempt to produce reliable and valid results.

The scientific method can be likened to an association of techniques rather than an exact formula; therefore, we expand the steps as a means to be more specific and relevant for research in education and the social sciences. As seen in Figure 1.1, these steps include the following: (a) identify a research problem, (b) establish the theoretical framework, (c) indicate the purpose and research questions (or hypotheses), (d) develop the methodology, (e) collect the data, (f) analyze and interpret the data, and (g) report the results. This book targets the critical component of the scientific method, referred to in Figure 1.1 as *Design the Study*, which is the point in the process when the appropriate research design is selected. We do not focus on prior aspects of the scientific method or any steps that come after the Design the Study step, including procedures for conducting literature

reviews, developing research questions, or discussions on the nature of knowledge, epistemology, ontology, and worldviews. Specifically, this book focuses on the conceptualization, selection, and application of common research designs in the field of education and the social and behavioral sciences.

Again, although the general premise is the same, the scientific method is known to slightly vary from each field of inquiry (and type of method). The technique presented here may not exactly follow the logic required for research using qualitative methods; however, the conceptualization of research designs remains the same. We refer the reader to Jaccard and Jacoby (2010) for a review on the various scientific approaches associated with qualitative methods, such as emergent- and discovery-oriented frameworks.

Figure 1.1 The Scientific Method



VALIDITY AND RESEARCH DESIGNS ♦

The overarching goal of research is to reach valid outcomes based upon the appropriate application of the scientific method. In reference to

research designs, *validity* is defined as the extent to which the outcome accurately answers the stated research questions of the study. Validity is a complex construct and takes on many different forms, operates on a continuum, and theoretically can be considered multidimensional. In other words, the outcome of most studies cannot typically be dichotomized as valid or not valid. Validity also has a place in psychometrics (i.e., the theories and techniques associated with educational and psychological measurements), and it is generally known as test validity.

The validity of a measurement tool simply means that it measures what it is developed to measure. The focus within this book is the validity related to research designs, *not* test validity (for more information related to test validity, reliability, and measurement, see DeVellis [2011] and Viswanathan [2005]). Although securing validity is critical at the design stage, it should be a consideration throughout the general steps of the scientific method. The importance of securing “acceptable” levels of validity for research in quantitative methods cannot be overstated. However, aspects of validity have also been addressed for qualitative methods. Validity and the qualitative method include focusing in on the *trustworthiness* of the data, such as Lincoln and Guba’s (2013) evaluation criteria, as well as the rigor and quality of the data collection procedures (see also Golafshani, 2003; Loh, 2013; Williams & Morrow, 2009). Additionally, the concept of external validity can have a place in qualitative methods as well. We refer the reader to Chenail (2010) for a review on nonprobabilistic approaches to aspects of generalizability for qualitative methods.

In the following sections, we summarize four types of validity related to research designs for quantitative methods: internal, external, construct, and statistical conclusion validity. Originally, the concepts of internal, external, construct, and statistical conclusion validity were all conceptualized for the application and development of experimental and quasi-experimental research (Campbell, 1957; Cook & Campbell, 1979). Since that time, many researchers, books, and Internet references have attempted to classify and order these types of validity very differently in accordance with nonexperimental research, as well as within different disciplines (e.g., epidemiology).

With minor additions, we organize and present the types of validity primarily based on Cook and Campbell’s (1979) original work, along with Shadish, Cook, and Campbell’s (2002) composition. Any condition that compromises the validity related to a research design is known as a *threat* (i.e., confounding variables). All types of validity are applicable to experimental and quasi-experimental research; however, the conceptualization of internal validity (by definition) does *not* apply to nonexperimental research,

including survey and observational (correlational) approaches. Another form of validity—statistical conclusion validity—applies to all research within quantitative methods and refers to the role of statistical analyses and its relation to research design.

Independent and Dependent Variables

In simple terms, the independent variable (IV) is the variable that is manipulated (i.e., controlled) by the researcher as a means to test its impact on the dependent variable, otherwise known as the *treatment effect*. In the classical experimental study, the IV is the treatment, program, or intervention. For example, in a psychology-based study, the IV can be a cognitive-behavioral intervention; the intervention is manipulated by the researcher, who controls the frequency and intensity of the therapy on the subject. In a pharmaceutical study, the IV would typically be a treatment pill, and in agriculture the treatment often is fertilizer. In regard to experimental research, the IVs are always manipulated (controlled) based on the appropriate theoretical tenets that posit the association between the IV and the dependent variable.

Statistical software packages (e.g., SPSS) refer to the IV differently. For instance, the IV for the analysis of variance (ANOVA) in SPSS is the “break-down” variable and is called a *factor*. The IV is represented as levels in the analysis (i.e., the treatment group is Level 1, and the control group is Level 2). For nonexperimental research that uses regression analysis, the IV is referred to as the *predictor variable*. In research that applies control in the form of statistical procedures to variables that were not or cannot be manipulated, the IVs are sometimes referred to as *quasi-* or *alternate independent variables*. These variables are typically demographic variables, such as gender, ethnicity, or socioeconomic status. As a reminder, in nonexperimental research the IV (or predictor) is not manipulated whether it is a categorical variable such as hair color or a continuous variable such as intelligence. The only form of control that is exhibited on these types of variables is that of statistical procedures. Manipulation and elimination do not apply (see *types of control* later in the chapter).

The dependent variable (DV) is simply the outcome variable, and its variability is a function of IV and its impact on it (i.e., treatment effect). For example, what is the impact of the cognitive-behavioral intervention on psychological well-being? In this research question, the DV is psychological well-being. In regard to nonexperimental research, the IVs are not manipulated, and the IVs are referred to as predictors and the DVs are criterion

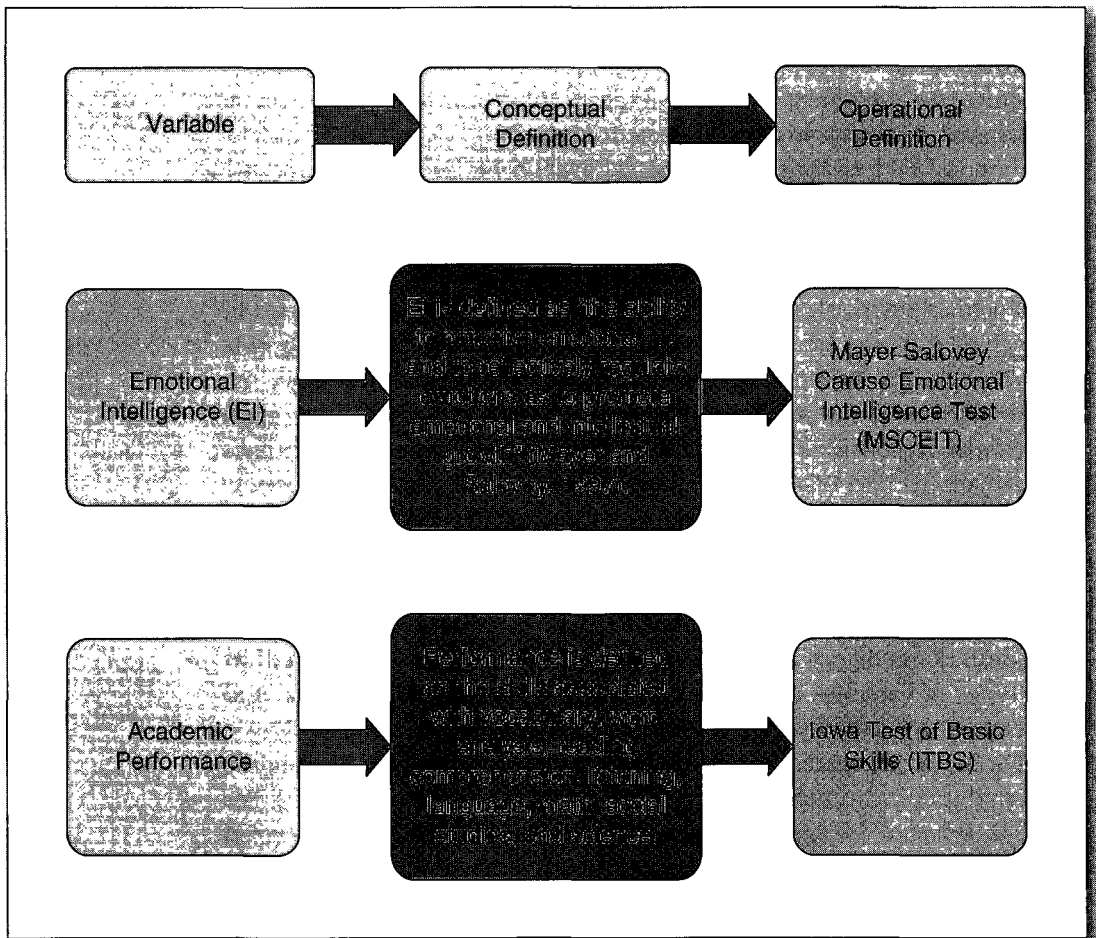
variables. During the development of research questions, it is critical to first define the DV conceptually, then define it operationally.

A **conceptual definition** is a critical element to the research process and involves scientifically defining the construct so it can be systematically measured. The conceptual definition is considered to be the (scientific) textbook definition. The construct must then be operationally defined to model the conceptual definition.

An **operational definition** is the actual method, tool, or technique that indicates how the construct will be measured (see Figure 1.2).

Consider the following example research question: What is the relationship between *Emotional Intelligence* and conventional *Academic Performance*?

Figure 1.2 Conceptual and Operational Definitions



Internal Validity

Internal validity is the extent to which the outcome was based on the independent variable (i.e., the treatment), as opposed to extraneous or unaccounted-for variables. Specifically, internal validity has to do with causal inferences—hence, the reason why it does not apply to nonexperimental research. The goal of nonexperimental research is to describe phenomena or to explain or predict the relationship between variables, not to infer causation (although there are circumstances when cause and effect can be inferred from nonexperimental research, and this is discussed later in this book). The identification of any explanation that could be responsible for an outcome (effect) outside of the independent variable (cause) is considered to be a threat. The most common threats to internal validity seen in education and the social and behavioral sciences are detailed in Table 1.1. It should be noted that many texts do not identify *sequencing effects* in the common lists of threats; however, it is placed here, as it is a primary threat in repeated-measures approaches.

Table 1.1 Threats to Internal Validity

<i>Threat</i>	<i>Explanation</i>
History	Any event that occurs during the time of the treatment and the posttest that could affect the outcome (e.g., natural life events such as a death in the family, change in job, or moving)
Maturation	The natural process of changing, growing, and learning over time
Testing	The effects of practice familiarity in taking the same test more than once (e.g., the participant who takes the same math achievement test twice in the pretest and posttest measures may improve performance simply because of the familiarity with the test)
Instrumentation	The change in a measuring instrument over time (i.e., some instruments undergo revisions)
Statistical regression	The tendency for any extreme score to regress toward the average (i.e., regression toward the mean is a statistical phenomenon that any extreme scores, high or low, eventually regress or revert to the average)
Selection bias	Also known as <i>selection effect</i> ; results when researchers do not use a systematic assignment technique (e.g., random assignment) to assign participants to conditions and is the largest threat to internal validity in quasi-experimental research

(Continued)

Table 1.1 (Continued)

<i>Threat</i>	<i>Explanation</i>
Attrition	The loss of participants during the term of the experiment (also known as <i>drop-out</i> or <i>subject mortality</i>)
Combination of selection and other treatments	For designs that include more than one group—any one of the threats to internal validity can affect one of the groups in the study as opposed to the other (e.g., the participants in one condition may have been exposed to a stressful event not related to the experiment, but this event does not affect the other condition)
Diffusion	The inadvertent application of the treatment to the control group (e.g., in educational settings, teachers may use aspects of the math intervention in the control group that are supposed to be delivered only to the control condition)
Special treatment	Special attention to the control group, with the changes attributed only to the attention (i.e., placebo effect)
Sequencing effects –Order effects –Carryover effects	Related to within-subject (repeated-measures) approaches and also known as <i>multiple-treatment interference</i> , <i>fatigue effects</i> , and <i>practice effects</i> ; can be separated into <i>order effects</i> (i.e., the order in which participants receive the treatment can affect the results) and <i>carryover effects</i> (i.e., performance in one condition affects performance in another condition)

External Validity

External validity is the extent to which the results can be generalized to the relevant populations, settings, treatments, or outcomes. Generally speaking, external validity can be secured if a true probability sampling technique (e.g., random selection) is used, although logistically this is often extremely difficult. Therefore, it is feasible that cause and effect can be established via the application of a sound experiment, but the findings may not generalize to the appropriate population or settings. As seen in Table 1.2, the primary threats to external validity are detailed and primarily slanted toward the examinations of causal relationships. However, issues pertaining to external validity should be considered for nonexperimental research. The most obvious threat to external validity for survey approaches (a form of nonexperimental research), for example, would be *sample characteristics*, sometimes referred to as *sampling bias*.

Table 1.2 Threats to External Validity

<i>Threat</i>	<i>Explanation</i>
Sample characteristics	The extent to which the sample (i.e., unit) represents the population from which it is drawn (i.e., for a sample to represent a population, the researcher must employ random selection and the appropriate sampling procedure and power analysis)
Stimulus characteristics and settings	The unique factors involved in providing the treatment or intervention, such as the setting and researchers (i.e., it is difficult to replicate contrived laboratory conditions to real-life scenarios)
Treatment variations	Variations in the same treatment or the combination of multiple or partial treatments that account for different results
Outcome variations	Observing the effect of one type of outcome differs when alternate outcomes are observed
Context-dependent mediation	Mediating variables related to outcomes differ between contexts or settings

Construct Validity

Construct validity refers to the extent a generalization can be made from the operationalization (i.e., the scientific measurement) of the theoretical construct back to the conceptual basis responsible for the change in the outcome. Again, although the list of threats to construct validity seen in Table 1.3 are defined to imply issues regarding cause-effect relations, the premise of construct validity should apply to all types of research. Some authors categorize some of these threats as *social threats* to internal validity, and some authors simply categorize some of the threats listed in Table 1.3 as threats to internal validity. The categorization of these threats can be debated, but the premise of the threats to validity cannot be argued (i.e., a violation of construct validity affects the overall validity of the study in the same way as a violation of internal validity).

Statistical Conclusion Validity

Statistical conclusion validity is the extent to which the statistical covariation (relationship) between the treatment and the outcome is accurate. Specifically, the statistical inferences regarding statistical conclusion validity

Table 1.3 Threats to Construct Validity

<i>Threat</i>	<i>Explanation</i>
Attention and contact with participants	Similar to <i>special treatment</i> ; the level of attention (differentiated attention) from the experimenter varies between the groups (e.g., the researcher spends more time with Group 1 than Group 2, and the differences observed in the outcome can be explained by the increased amount of attention and not due to the intervention)
Single operations and narrow stimulus sampling	The impact the researcher has on the development and implementation of the treatment (i.e., researchers deliver treatments differently based on experiences and expertise; therefore, it is difficult to measure the impact the researcher has on the treatment itself)
Experimenter expectancies	The researchers' expectancies, beliefs, and biases about the results (e.g., if a researcher strongly believes anxiety reduces test performance, then the interaction between the researcher and the participant may influence the outcome because the delivery of instructions and adherence to protocols may change)
Cues of the experimental situation	Sources of influence conveyed to prospective participants (e.g., rumors, information passed along from previous participants)
Novelty effects	The novelty of being in a new or innovative context
Inadequate explication of constructs	The construct under investigation is not appropriately defined conceptually, leading to inadequate measurement (i.e., operationalization)
Construct confounding	Multiple constructs not clearly identified and accounted for operationally
Mono-operation bias	An operationalization (i.e., measurement) does not appropriately represent the construct under investigation, leading to measuring unintended constructs
Mono-method bias	All measurement techniques are the same as a means to measure the construct under investigation
Confounding constructs with levels of constructs	All the levels of a construct are not fully accounted for through the appropriate measurement and reporting tools
Treatment sensitive factorial structure	The interpretation and structure of a measure change as a result of the treatment
Reactivity to assessment	The participants' awareness of being studied may influence the outcome; also known as <i>acquiescence bias</i> , <i>social desirability</i> , and the <i>Hawthorne</i> or <i>observer effect</i> ; also an unnatural reaction to any particular form of assessment (i.e., when participants know they are being assessed, the assessment is considered obtrusive and may alter outcome measures other than what they would naturally)

<i>Threat</i>	<i>Explanation</i>
Test sensitization	Also known as <i>pretest sensitization</i> ; the sensitization to the intervention when participants are pretested (e.g., participants are pretested on perceptions of persuasive speeches and are then shown a movie on a persuasive speech; the pretest may influence how they view the speech)
Timing of measurement	The point in time the assessments are administered (i.e., unknown changes may occur, and the different timing of assessments may reveal different results)
Compensatory equalization	When participants in one condition receive more desirable services or compensation compared to that of another condition (thus, constituents may provide enhanced services or goods to the condition not receiving the benefits)
Compensatory rivalry	When participants in the control condition make a concerted effort to make improvements or changes in line with the treatment condition
Resentful demoralization	When participants become resentful or demoralized when they perceive they are receiving a less desirable treatment compared to that of another condition

has to do with the ability with which one can detect the relationship between the treatment and outcome, as well as determine the strength of the relationship between the two. As seen in Table 1.4, the most notable threats to statistical conclusion validity are outlined. Violating a threat to statistical conclusion validity typically will result in the overestimation or underestimation of the relationship between the treatment and outcome in experimental research. A violation can also result in the overestimation or underestimation of the explained or predicted relationships between variables as seen in nonexperimental research.

Table 1.4 Threats to Statistical Conclusion Validity

<i>Threat</i>	<i>Explanation</i>
Low statistical power	Power is the extent to which the results of an analysis accurately reveal a statistically significant difference between groups (or cases) when a statistical difference truly exists.
Assumption violation of statistical tests	Violating the assumptions (depending on the extent of the violation) of statistical tests can lead to overestimation or underestimation of practical and statistical significance of an outcome.

(Continued)

Table 1.4 (Continued)

<i>Threat</i>	<i>Explanation</i>
Error rate problem	Statistical significance can be artificially inflated when performing multiple pairwise tests; it is also referred to as <i>family-wise error rate</i> (i.e., the probability of making a Type I error when performing multiple pairwise analyses).
Restriction of range	A lack of variability between variables weakens the relationship and lowers statistical power.
Extraneous variance in the experimental setting	Variations within the experimental setting (e.g., temperature) may inflate error.
Inaccurate effect size estimation	Some statistical analyses can overestimate or underestimate the size of an effect.
Variability in the procedures	Also referred to as <i>unreliability of treatment implementation</i> , the variations in the application of an intervention may affect the outcome (i.e., a nonstandardized approach will create variability in the outcome that is not attributable to the treatment, but rather to the application of the treatment).
Subject heterogeneity	The variability of participant demographics (e.g., age, race, ethnicity, background) may create unaccounted-for variations in the findings.
Unreliability of the measures	Measures maintain certain levels of validity and reliability (pertaining to psychometric principles), and lack of reliability causes inconsistency in measurement.
Multiple comparisons and error rates	The use of multiple dependent variables across conditions and multiple statistical analyses creates greater opportunities for error variance.

The reader is referred to the following books and article for an in-depth review of issues related to validity in research:

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.

Shadish, W. R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods, 15*, 3–17.

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

The overarching objective of a research design is to provide a framework from which specific research questions or hypotheses can be answered while using the scientific method. The concept of a research design and its structure is, at face value, rather simplistic. However, complexities arise when researchers apply research designs within social science paradigms. These include, but are not limited to, logistical issues, lack of control over certain variables, psychometric issues, and theoretical frameworks that are not well developed. In addition, with regard to statistical conclusion validity, a researcher can apply sound principles of scientific inquiry while applying an appropriate research design but may compromise the findings with inappropriate data collection strategies, faulty or “bad” data, or misdirected statistical analyses. Shadish and colleagues (2002) emphasized the importance of structural design features and that researchers should focus on the theory of design logic as the most important feature in determining valid outcomes (or testing causal propositions). The logic of research designs is ultimately embedded within the scientific method, and applying the principles of sound scientific inquiry within this phase is of the utmost importance and the primary focus of this guide.

Control

Control is an important element to securing the validity of research designs within quantitative methods (i.e., experimental, quasi-experimental, and nonexperimental research). However, within qualitative methods, behavior is generally studied as it occurs naturally with no manipulation or control. Control refers to the concept of holding variables constant or systematically varying the conditions of variables based on theoretical considerations as a means to minimize the influence of unwanted variables (i.e., extraneous variables). Control can be applied actively within quantitative methods through (a) manipulation, (b) elimination, (c) inclusion, (d) group or condition assignment, or (e) statistical procedures.

Manipulation. Manipulation is applied by manipulating (i.e., controlling) the independent variable(s). For example, a researcher can manipulate a behavioral intervention by systematically applying and removing the intervention or by controlling the frequency and duration of the application (see section on independent variables).

Elimination. Elimination is conducted when a researcher holds a variable or converts it to a constant. If, for example, a researcher ensures the temperature in a lab is set exactly to 76° Fahrenheit for both conditions in a biofeedback study, then the variable of temperature is eliminated as a factor because it is held as a constant.

Inclusion. Inclusion refers to the addition of an extraneous variable into the design to test its affect on the outcome (i.e., dependent variable). For example, a researcher can include both males and females into a factorial design to examine the independent effects gender has on the outcome. Inclusion can also refer to the addition of a control or comparison group within the research design.

Group assignment. Group assignment is another major form of control (see more on group and condition assignments later). For the between-subjects approach, a researcher can exercise control through random assignment, using a matching technique, or applying a cutoff score as means to assign participants to conditions. For the repeated-measures approach, control is exhibited when the researcher employs the technique of counterbalancing to variably expose each group or individual to all the levels of the independent variable.

Statistical procedures. Statistical procedures are exhibited on variables, for example, by systematically deleting, combining, or not including cases and/or variables (i.e., removing outliers) within the analysis. This is part of the data-screening process as well. As illustrated in Table 1.5, all of the major forms of control can be applied in the application of designs for experimental and quasi-experimental research. The only form of control that can be applied to nonexperimental research is statistical control.

Table 1.5 Control Techniques for Experimental, Quasi-Experimental, and Nonexperimental Research

<i>Type of Control</i>	<i>Experimental and Quasi-Experimental Research</i>	<i>Nonexperimental Research</i>
Manipulation	Yes	—
Elimination	Yes	—
Inclusion	Yes	—
Group or condition assignment	Yes	—
Statistical procedures	Yes	Yes

DESIGN NOTATIONS ♦

Design notations are the symbols used to diagrammatically illustrate the process of a research design (see Table 1.6). Within the design, time moves from left to right of the design structure. We used the design notations presented here in each research design covered. The notations presented in this book are based on Campbell and Stanley's (1963) work.

Observation (O). Observation, also known as *measurement*, is symbolized by an "O." The O can refer to a single measure of the dependent variable or multiple measures ($O_1, O_2 \dots O_n$).

Treatment (X). Treatment, also known as *intervention* or *program* (i.e., the treatment is technically the independent variable and also referred to as a factor), is symbolized with an "X." A control group typically does not receive the treatment and is designated as "-" in its place.

Factor (A, B . . . Z). Multiple treatments (factors) used in a design are designated as "X_A" and "X_B" and can go as far up the alphabet as there are factors.

Table 1.6 Design Notations

Design Notation	Design Element
O	Observation
X	Treatment
A, B	Factor

ASSIGNMENT TECHNIQUES ♦

In quantitative methods, each group in a research design has its own line within the structure of the diagram (see Table 1.7). One line equates to one group, two lines equate to two groups, and so on. The assignment of a group is usually the first design notation listed in the line structure.

Random assignment (R) Participants are randomly assigned to each condition to theoretically ensure group equivalency. Logistically, as seen in Figure 1.3, *stratified* random assignment (R_s), sometimes referred to as *blocking*, is used to ensure that

Table 1.7 Group Assignment Design Notations

Design Notation	Assignment
R	Random
NR	Nonrandom
C	Cutoff score
M	Matched

the subjects are balanced within predetermined stratum blocks or strata (e.g., age, ethnicity) and then randomly assigned to conditions. See Imgen and Rubin (2015) for more on classical random-assignment approaches, such as Bernoulli trials, completely randomized, stratified, and paired-randomized experiments.

Nonrandom assignment (NR). Participants are assigned to each condition by a matter of convenience or necessity because random assignment is neither an option nor required (nonequivalent groups).

Cutoff score (C). A cutoff score (criterion) is used to assign participants to groups within regression-discontinuity approaches. To create a cutoff criterion, a single pretest continuous distribution is determined and then a division in the data (i.e., cutoff) is made that determines the assignment of participants to conditions.

Matched (M). Matching is a technique used by researchers to match participants on the basis of some extraneous variable that is related to the dependent variable. When this technique is used to assign participants to conditions, some researchers refer to these as match-group designs, but this is not entirely accurate. It is the assignment technique that changes, but the design remains the same.

Matched pairs. For application in any research design indicated in the between-subjects approach, the researcher can (a) match participants in pairs based on certain criteria (e.g., IQ score), then randomly assign each member of the pair to conditions in order to ensure group equivalency (experimental design), and designate this as M_r or (b) match participants based on certain criteria without random assignment to a specific group (quasi-experimental design), then designate this as M_{NR} . For more on matched pairs, see Shadish et al. (2002, p. 118).

Matched grouping. For application in observational approaches, as well as the ex post facto (i.e., after the fact) design, the researcher manually matches participants in groups (M_A) as a means to establish control over the variables of interest. This is conducted because the independent [treatment] variable has already occurred and is not manipulated; therefore, various levels of alternate independent variables (e.g., age, gender) can be statistically manipulated and used as a means to assign individuals to conditions (see more on ex post facto designs later in this guide). This is a form of statistical procedures control often used in epidemiology studies.

Counterbalancing. Counterbalancing is a technique used only in repeated-measures approaches to control for *sequencing effects*. Researchers use counterbalancing to variably expose each group or individual to all the treatments or various treatment levels. The most common form of counterbalancing is conducted at the group level (each group is exposed to the treatment at different sequences). However, counterbalancing can be randomized (sequence is randomly determined for each participant), intrasubject (participants are exposed to more than one sequence, usually in one order, then reversed), complete (every possible sequence is offered), or incomplete (not every sequence is provided because it would require too many conditions, as seen later in the Latin-square design).

The reader is referred to the following article and book for an in-depth review of topics related to group assignment:

Cook, T. D., & Steiner, P. M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pre-test measures of outcome, of unreliable measurement, and of mode of data analysis. *Psychological Methods, 15*(1), 56–68.

Rubin, D. B. (2006). *Matched sampling for causal effects*. Cambridge, England: Cambridge University Press.

Figure 1.3 Example of a Stratified Random-Assignment Technique

Sample of Subjects With GPAs Ranging From 2.0 to 4.0 (N = 52)				
	Subjects With a GPA of 2.0 to 2.5 (n = 14)	Subjects With a GPA of 2.6 to 3.0 (n = 12)	Subjects With a GPA of 3.1 to 3.5 (n = 16)	Subjects With a GPA of 3.6 to 4.0 (n = 10)
	↓	↓	↓	↓
1 Treatment (x)	n = 7	n = 6	n = 8	n = 5
2 Control (-)	n = 7	n = 6	n = 8	n = 5

Note: This is an example of a two-group design (one treatment and one control group), and the pool of subjects is separated into strata based on grade point average (GPA; i.e., the stratification variable) and then randomly assigned to conditions. Some researchers recommend using this technique when $N < 100$ (Lachin, Matts, & Wei, 1988).

♦ COMPARISON AND CONTROL GROUPS

The group that does not receive the actual treatment, or intervention, is typically designated as the control group. Control groups fall under the *group or condition assignment* aspect of control. Control groups are comparison groups and are primarily used to address threats to internal validity such as history, maturation, selection, and testing. A *comparison group* refers to the group or groups that are not part of the primary focus of the investigation but allow the researcher to draw certain conclusions and strengthen aspects of internal validity. There are several distinctions and variations of the control group that should be clarified.

Control group. The control group, also known as the *no-contact control*, receives no treatment and no interaction.

Attention control group. The attention control group, also known as the *attention-placebo*, receives attention in the form of a pseudo-intervention to control for reactivity to assessment (i.e., the participant's awareness of being studied may influence the outcome).

Nonrandomly assigned control group. The nonrandomly assigned control is used when a no-treatment control group cannot be created through random assignment.

Wait-list control group. The wait-list control group is withheld from the treatment for a certain period of time, then the treatment is provided. The time in which the treatment is provided is based on theoretical tenets and on the pretest and posttest assessment of the original treatment group.

Historical control group. Historical control is a control group that is chosen from a group of participants who were observed at some time in the past or for whom data are available through archival records, sometimes referred to as *cohort* controls (i.e., a homogenous successive group) and useful in quasi-experimental research.

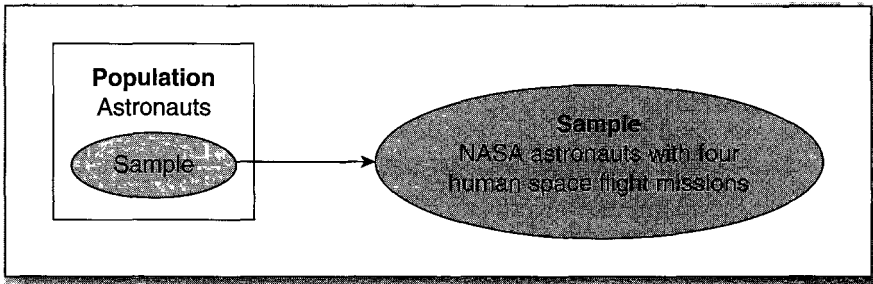
♦ SAMPLING STRATEGIES

A major element to the logic of design extends to sampling strategies. When developing quantitative, qualitative, and mixed methods studies, it is

important to identify the individuals (or extant databases) from whom you plan to collect data. To start, the *unit of analysis* must be indicated. The unit of analysis is the level or distinction of an entity that will be the focus of the study. Most commonly, in social science research, the unit of analysis is at the individual or group level, but it can also be at the programmatic level (e.g., institution or state level).

There are instances when researchers identify units nested within an aggregated group (e.g., a portion of students within a classroom) and refer to this as *nested designs* or models. It should be noted that examining nested units is not a unique design, but rather a form of a sampling strategy, and the relevant aspects of statistical conclusion validity should be accounted for (e.g., independence assumptions). After identifying the unit, the next step is to identify the *population* (assuming the individual or group is the unit of analysis), which is the group of individuals who share similar characteristics (e.g., all astronauts). Logistically, it is impossible in most circumstances to collect data from an entire population; therefore, as illustrated in Figure 1.4, a *sample* (or subset) from the population is identified (e.g., astronauts who have completed a minimum of four human space-flight missions and work for NASA).

Figure 1.4 Example of a Sample Extracted From a Population



The goal often, but not always, is to eventually generalize the finding to the entire population. There are two major types of sampling strategies, probability and nonprobability sampling. In experimental, quasi-experimental, and nonexperimental (survey and observational) research, the focus should be on probability sampling (identifying and selecting individuals who are considered representative of the population). Many researchers also suggest that some form of probability sampling for observational (correlational) approaches (predictive designs) must be employed—otherwise the statistical outcomes cannot be generalizable. When it is not logistically possible to use probability sampling, or as

seen in qualitative methods not necessary, some researchers use nonprobability sampling techniques (i.e., the researcher selects participants on a specific criterion and/or based on availability). The following list includes the major types of probability and nonprobability sampling techniques.

Probability Sampling Techniques

Simple random sampling. Every individual within the population has an equal chance of being selected.

Cluster sampling. Also known as *area sampling*, this allows the researcher to divide the population into clusters (based on regions) and then randomly select from the clusters.

Stratified sampling. The researcher divides the population into homogeneous subgroups (e.g., based on age) and then randomly selects participants from each subgroup.

Systematic sampling. Once the size of the sample is identified, the researcher selects every *n*th individual (e.g., every third person on the list of participants is selected) until the desired sample size is fulfilled.

Multistage sampling. The researcher combines any of the probability sampling techniques as a means to randomly select individuals from the population.

Nonprobability Sampling Techniques

Convenience sampling. Sometimes referred to as *baphazard* or *accidental sampling*, the investigator selects individuals because they are available and willing to participate.

Purposive sampling. The researcher selects individuals to participate based on a specific need or purpose (i.e., based on the research objective, design, and target population); this is most commonly used for qualitative methods (see Patton, 2002). The most common form of purposeful sampling is *criterion sampling* (i.e., seeking participants who meet a specific criterion). Variations of purposive sampling include *theory-guided*, *snowball*, *expert*, and *heterogeneity sampling*. *Theoretical sampling* is a type of purposive sampling used in grounded-theory approaches. We refer the reader to Palinkas et al. (2014) for a review of recommendations on how to combine various sampling strategies for the qualitative and mixed methods.

The reader is referred to the following book for an in-depth review of a topic related to sampling strategies for quantitative and qualitative methods:

Levy, P. S., & Lemeshow, S. (2009). *Sampling of populations: Methods and applications* (4th ed.). New York, NY: John Wiley & Sons.

Now that we covered a majority of the relevant aspects to research design, which is the “Design the Study” phase of the scientific method, we now present some steps that will help researchers select the most appropriate design. In the later chapters, we present a multitude of research designs used in quantitative, qualitative, and mixed methods. Therefore, it is important to review and understand the applications of these designs while regularly returning to this chapter to review the critical elements of design control and types of validity, for example. Let’s now examine the role of the research question.

RESEARCH QUESTIONS ♦

Simply put, the primary research question sets the foundation and drives the decision of the application of the most appropriate research design. However, there are several terms related to research questions that should be distinguished. First, in general, studies will include an overarching observation deemed worthy of research. The “observation” is a general statement regarding the area of interest and identifies the area of need or concern.

Based on the initial observation, specific variables lead the researchers to the appropriate review of the literature and a theoretical framework is typically established. The purpose statement is then used to clarify the focus of the study, and finally, the primary research question ensues. Research studies can also include hypotheses or research objectives. Many qualitative studies include research aims as opposed to research questions. In quantitative methods (this includes mixed methods), the research question (hypotheses and objectives) determines (a) the population (and sample) to be investigated, (b) the context, (c) the variables to be operationalized, and (d) the research design to be employed.

Types of Inquiry

There are several ways to form a testable research inquiry. For qualitative methods, these can be posed as research questions, aims, or objectives

while identifying the central phenomenon to be explored. For the application of quantitative methods, researchers can use questions and objectives as well, but also can use hypothesis. Hypotheses are simply predictions the researcher posits as to the direction a relationship will manifest between two or more variables. A hypothesis is purely statistical terminology that is thus tested with statistics. At the heart of every statistical analysis is the null hypothesis. For example, a basic *t* test is used to examine the mean differences between two groups. The null hypothesis for the *t* test is that no differences exist between the two groups. The researcher then collects data from the two groups, states an alternate hypothesis to the null, and then analyzes the data with the *t* test to either reject or accept that null. And in the process, the hypothesis is confirmed or disconfirmed.

Research questions for the quantitative method are still tested in the same manner but are just presented in a different fashion. Creswell's (2014) composition presented three major types of research questions and scripts to be applied to aid in the development of these questions. The three types are the following:

Descriptive. The descriptive question indicates the participants and at least one variable to be investigated. An example could be "What are the anxiety levels of students in the math class?" In this example, the variable to be measured is anxiety levels, and the participants are students in a math class.

Relational. A relationship question includes at least two variables and the participants from which the data should be collected. For example, "What is the relationship between pretest anxiety and test scores for students taking college entrance exams?" The two variables are anxiety and test scores.

Comparison. A comparison question indicates at least two distinct groups and at least one variable that can be measured between the two groups. For example, "How do males compare to females in terms of their pretest anxiety and test scores on college entrance exams?"

Research questions for the qualitative method are classified as central and subquestions. It is recommended to begin qualitative research questions with open-ended verbs such as *what* or *how* to convey the emerging aspect reflective of the qualitative method.

Central. The central research question is a broad statement of inquiry focused on the exploration of the central or primary phenomenon of focus. For example, a central research question for an ethnographic

approach could be “How do Latin-American immigrant children transition into the English-speaking school system?”

Subquestion. The subquestions follow the central question and narrow the focus. The subquestions are a starting point to the development of the qualitative data collection procedures (e.g., interview or focus group questions). Follow-up subquestions, for example, could be “What are the experiences of Latin-American students in the school?” and “How are these experiences reflected at home with their family?”

A flowchart and examples follow that will assist researchers in determining the most appropriate design based on the primary research question of the study. Recall from the Preface the chart that indicated the levels related to determining a design for quantitative and qualitative methods (Method, Research, Approach, and Design). The research question can be broken down, using this chart to determine the most appropriate design.

QUANTITATIVE	
Level	Explanation
METHOD ₁	The <i>method</i> provides the theoretical, philosophical, and data analytic stance (e.g., a quantitative method ₁).
▼	▼
RESEARCH ₂	<i>Research</i> refers to the systematic process of control (e.g., group assignment, selection, and data collection techniques). Research can be experimental, quasi-experimental, or nonexperimental (e.g., a quantitative method ₁ and experimental research ₂).
▼	▼
APPROACH ₃	The <i>approach</i> is the first step to creating structure to the design, and it details (a) a theoretical model of how the data will be collected, and (b) if one case, one group, or multiple groups will be associated with the process (e.g., a quantitative method ₁ , experimental research ₂ with a between-subjects approach ₃).
▼	▼
DESIGN ₄	The <i>design</i> is the actual structure or framework that indicates (a) the time frame(s) in which data will be collected, (b) when the treatment will be implemented (or not), and (c) the exact number of groups that will be involved (e.g., a quantitative method ₁ , experimental research ₂ with a between-subjects approach ₃ and a pre- and posttest control group design ₄).

QUALITATIVE	
METHOD ₁	The <i>method</i> provides the theoretical, philosophical, and data analytic stance (e.g., a qualitative method ₁).
▼	▼
RESEARCH ₂	<i>Research</i> for the qualitative method is nonexperimental (e.g., a qualitative method ₁ and nonexperimental research ₂).
▼	▼
PERSPECTIVE ₃	The <i>perspective</i> is the first step to creating structure to the design, and it details the theoretical perspective (or lens) of how the researcher(s) will approach the study (e.g., a qualitative method ₁ , nonexperimental research ₂ with an ethnographic perspective ₃).
▼	▼
DESIGN ₄	The <i>design</i> is the actual structure that indicates (a) if one case, one group, or multiple groups will be associated with the process, and (b) when the data will be analyzed (e.g., a qualitative method ₁ , nonexperimental research ₂ with an ethnographic ₃ and a case study design ₄).

Type of Research Question	
Level	Question
METHOD ₁	Quantitative or qualitative
▼	▼
RESEARCH ₂	Experimental, quasi-experimental, or nonexperimental
▼	▼
APPROACH ₃	Quantitative or qualitative methodological variant
▼	▼
DESIGN ₄	Any design variant found under the quantitative or qualitative method

Example 1.1

Descriptive	
Level	What are the levels of perceived anxiety students experience prior to testing?
METHOD ₁	Quantitative
▼	▼
RESEARCH ₂	Nonexperimental

<i>Descriptive</i>	
▼ APPROACH ₃	▼ Survey
▼ DESIGN ₄	▼ Cross-sectional

Note: Perceived anxiety is the only variable in this question that requires operationalization. It is likely that a cross-sectional design will suffice, but if time allows for it, a longitudinal design can be employed.

Example 1.2

<i>Relational</i>	
Level	To what extent do levels of perceived anxiety predict performance on standardized testing?
METHOD ₁	Quantitative
▼ RESEARCH ₂	▼ Nonexperimental
▼ APPROACH ₃	▼ Observational
▼ DESIGN ₄	▼ Predictive

Note: The variables in this question are anxiety and test performance. This is a relational question that qualifies as an observational approach. The design can be explanatory, but if the data points are not collected at the same time (i.e., anxiety collected at Time Point 1 and then test performance at Time Point 2), then a predictive form of analysis can be used to reduce the data for further interpretation and discussion.

Example 1.3

<i>Comparison</i>	
Level	How do the groups differ between the high-anxiety and low-anxiety conditions in terms of test performance?
METHOD ₁	Quantitative
▼ RESEARCH ₂	▼ Experimental

(Continued)

Example 1.3 (Continued)

<i>Comparison</i>	
▼ APPROACH ₃	▼ Between-subjects
▼ DESIGN ₄	▼ 2-factor posttest

Note: The research question includes one outcome variable broken down into two levels (high and low anxiety). This would require two groups to examine the differences. If random assignment to conditions is employed, then the research is experimental and only a 2-factor posttest design can be employed. If enough participants are available, a third group can be included and considered a control group. If time is on the researcher's side, then a pretest can be included as well, but it is not necessary, particularly if random assignment to conditions is employed.

Example 1.4

<i>Comparison</i>	
Level	How do the groups differ when exposed to the high-anxiety and low-anxiety conditions in terms of test performance?
METHOD ₁	Quantitative
▼ RESEARCH ₂	▼ Experimental
▼ APPROACH ₃	▼ Within-subjects
▼ DESIGN ₄	▼ 2-factor crossover

Note: Similar to the previous example, there is one outcome (dependent) variable at two levels. However, if the researcher has access to only a small group of participants, then a within-subjects (repeated-measures) approach can be used. The participants would experience both conditions through the application of the 2-factor crossover design.

Example 1.5

<i>Central Question</i>	
Level	What are the experiences of parents who have children diagnosed with a pervasive developmental disorder (PDD)?
METHOD ₁	Qualitative
▼ RESEARCH ₂	▼ Nonexperimental

<i>Central Question</i>	
▼ PERSPECTIVE ₃	▼ Narrative
▼ DESIGN ₄	▼ Descriptive

Note: The central phenomenon is the experience of parents who have children with PDDs. In this example, the researcher is interested in using the narrative perspective as a means to simply provide storytelling to understand the phenomenon. The descriptive design further delineates the perspective that the goal is to provide the narrative of the life stories without providing a critique or assuming there are causes for the resulting phenomenon.

Example 1.6

<i>Central Question</i>	
Level	What are the instructional approaches used by instructors to deal with multicultural populations in graduate school?
METHOD ₁	Qualitative
▼ RESEARCH ₂	▼ Nonexperimental
▼ PERSPECTIVE ₃	▼ Ethnographic
▼ DESIGN ₄	▼ Realist

Note: The phenomenon to be explored is the instructional approaches for multicultural populations. The ethnographic perspective is adequate in that it will guide the researcher to further understand the point of view of participants from varied cultural backgrounds. The instructional approaches can be culled down for reporting as guided through the realist design.

Keep in mind the examples only reflect general guidelines. Often, researchers pose multiple research questions, which are considered spinoffs of the primary questions. Although this doesn't change the research design, it guides the type of analysis required to properly interpret the data. In summary, if the primary question is descriptive, then the research will be non-experimental, and a survey approach should be employed. If the primary question is comparative, then any approach and design that falls under the category of quasi-experimental or experimental research should be used. If the primary question is relational, then an observational approach and a

predictive or explanatory design should be applied. As a reminder, the application of the appropriate design relative to the primary research can vary depending on the specific research scenario and the field from which the examination is to be applied. The reader is referred to White (2009) for an in-depth review of the development of research questions for social scientists.

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. Explain from a technical viewpoint why it is important to distinguish a method, research, approach, and design. Next, briefly discuss how understanding each term individually in addition to how these terms interconnect is important for your understanding of the application of research designs.
2. Discuss the importance of validity and research design. Next, choose one type of validity (internal, external, construct, or statistical conclusion) and discuss its relevance to experimental, quasi-experimental, and nonexperimental research.

Exercise

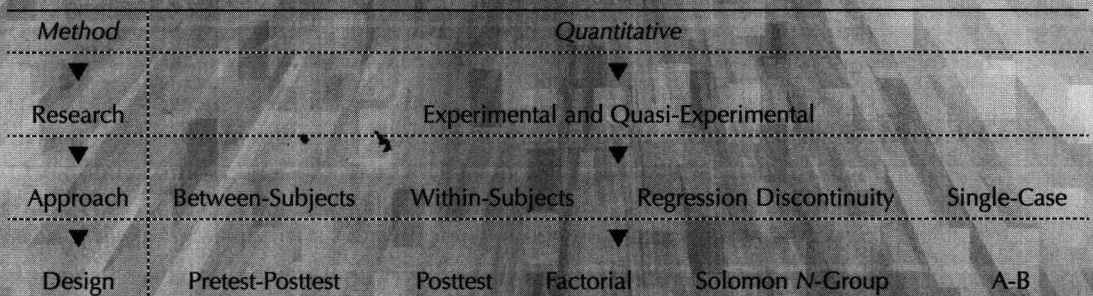
1. Define a sampling strategy.
2. Define the two major types of sampling strategies.
3. Identify a hypothetical population.
4. Identify the sample.
5. What type of sampling strategy will be used?
 - a. Why did you choose this type of strategy?
6. Based on the strategy, what type of sampling technique will be used to identify the sample?
 - a. Why did you choose this type of technique?

PART I

Quantitative Methods for Experimental and Quasi-Experimental Research

Part I includes four popular approaches to the quantitative method (experimental and quasi-experimental only), followed by some of the associated basic designs (accompanied by brief descriptions of published studies that used the design). Visit the companion website at study.sagepub.com/edmonds2e to access valuable instructor and student

Figure I.1 Quantitative Method Flowchart



Note: Quantitative methods for experimental and quasi-experimental research are shown here, followed by the approach and then the design.

resources. These resources include PowerPoint slides, discussion questions, class activities, SAGE journal articles, web resources, and online data sets.

Research in quantitative methods essentially refers to the application of the systematic steps of the scientific method, while using quantitative properties (i.e., numerical systems) to research the relationships or effects of specific variables. Measurement is the critical component of the quantitative method. Measurement reveals and illustrates the relationship between quantitatively derived variables. Variables within quantitative methods must be, first, conceptually defined (i.e., the scientific definition), then operationalized (i.e., determine the appropriate measurement tool based on the conceptual definition). Research in quantitative methods is typically referred to as a *deductive process* and iterative in nature. That is, based on the findings, a theory is supported (or not), expanded, or refined and further tested.

Researchers must employ the following steps when determining the appropriate quantitative research design. First, a measurable or testable research question (or hypothesis) must be formulated. The question must maintain the following qualities: (a) precision, (b) viability, and (c) relevance. The question must be *precise* and well formulated. The more precise, the easier it is to appropriately operationalize the variables of interest. The question must be *viable* in that it is logistically feasible or plausible to collect data on the variable(s) of interest. The question must also be *relevant* so that the result of the findings will maintain an appropriate level of practical and scientific meaning. The second step includes choosing the appropriate design based on the primary research question, the variables of interest, and logistical considerations. The researcher must also determine if randomization to conditions is possible or plausible. In addition, decisions must be made about how and where the data will be collected. The design will assist in determining when the data will be collected. The unit of analysis (i.e., individual, group, or program level), population, sample, and sampling procedures should be identified in this step. Third, the variables must be operationalized. And last, the data are collected following the format of the framework provided by the research design of choice.

♦ EXPERIMENTAL RESEARCH

Experimental research (sometimes referred to as *randomized experiments*) is considered to be the most powerful type of research in determining causation among variables. Cook and Campbell (1979) presented three conditions that must be met in order to establish cause and effect:

1. **Covariation** (the change in the cause must be related to the effect)
2. **Temporal precedence** (the cause must precede the effect)
3. **No plausible alternative explanations** (the cause must be the only explanation for the effect)

The essential features of experimental research are the sound application of the elements of control: (a) manipulation, (b) elimination, (c) inclusion, (d) group or condition assignment, or (e) statistical procedures. Random assignment (not to be confused with random selection) of participants to conditions (or random assignment of conditions to participants [counterbalancing] as seen in repeated-measures approaches) is a critical step, which allows for increased control (improved internal validity) and limits the impact of the confounding effects of variables that are not being studied.

The random assignment to each group (condition) theoretically ensures that the groups are “probabilistically” equivalent (controlling for selection bias), and any differences observed in the pretests (if collected) are considered due to chance. Therefore, if all threats to internal, external, construct, and statistical conclusion validity were secured at “adequate” levels (i.e., all plausible alternative explanations are accounted for), the differences observed in the posttest measures can be attributed fully to the experimental treatment (i.e., cause and effect can be established). Conceptually, a causal effect is defined as a comparison of outcomes derived from treatment and control conditions on a common set of units (e.g., school, person).

The strength of experimental research rests in the reduction of threats to internal validity. Many threats are controlled for through the application of random assignment of participants to conditions. *Random selection*, on the other hand, is related to sampling procedures and is a major factor in establishing external validity (i.e., generalizability of results). Randomly selecting a sample from a population would be conducted so that the sample would better represent the population. However, Lee and Rubin (2015) presented a statistical approach that allows researchers to draw data from existing data sets from experimental research and examine subgroups (post hoc subgroup analysis). Nonetheless, random assignment is related to design, and random selection is related to sampling procedures. Shadish, Cook, and Campbell (2002) introduced the term *generalized causal inference*. They posit that if a researcher follows the appropriate tenets of experimental design logic (e.g., includes the appropriate number of subjects, uses random selection and random assignment) and controls for threats of all types of validity (including test validity), then valid causal inferences can be determined along with the ability to generalize the causal link. This is truly

realized once multiple replications of the experiment are conducted and comparable results can be observed over time (replication being the operative word). Though, recently there have been concerns related to the reproducibility of experimental studies published in the field of psychology, for example (see Baker, 2015; Bohannon, 2015).

Reproducibility could be enhanced if the proper tenets of the scientific method are followed and the relevant aspects of validity are addressed (i.e., internal and construct). Researchers tend to gloss over these constructs and rarely report how they ensured the data to be valid, often assuming that a statistical analysis could be used to “fix” or overshadow the inherent problems of the data. Bad data is clearly the issue, which lends to a great computer science saying “Garbage in, garbage out.” To be more specific, taking the appropriate measures to ensure design and test validity, the data will be more “clean,” which results in fewer reporting errors in the statistical results. Although probability sampling (e.g., random selection) adds another logistical obstacle to experimental research, it should also be an emphasis along with the proper random assignment techniques.

Although this book is more dedicated to the application of research designs in the social and behavioral sciences, it is important to note the distinction between research designs in the health sciences to that of the social sciences. Experimental research in the health or medical sciences shares the same designs, although the terminology slightly differs, and the guidelines for reporting the data can be more stringent (e.g., see Schultz, Altman, & Moher, 2010, and Appendix H for guidelines and checklist). These guidelines are designed to enhance the quality of the application of the design, which in turn leads to enhanced reproducibility. The most common term used to express experimental research in the field of medicine is randomized control trials (RCT). RCT simply infers that subjects are randomly assigned to conditions. The most common of the RCT designs is the *parallel-group approach*, which is another term for the *between-subject approach* and is discussed in more detail in the following sections. RCTs can also be crossover and factorial designs and are designated under the within-subjects approach (repeated measures).

♦ QUASI-EXPERIMENTAL RESEARCH

The nonrandom assignment of participants to each condition allows for convenience when it is logistically not possible to use random assignment. Quasi-experimental research designs are also referred to as *field research* (i.e., research is conducted with an intact group in the field as opposed to

the lab), and they are also known as *nonequivalent designs* (i.e., participants are not randomly assigned to each condition; therefore, the groups are assumed nonequivalent). Hence, the major difference between experimental and quasi-experimental research designs is the level of control and assignment to conditions. The actual designs are structurally the same, but the analyses of the data are not. However, some of the basic pretest and posttest designs can be modified (e.g., addition of multiple observations or inclusion of comparison groups) in an attempt to compensate for lack of group equivalency. In the design structure, a dashed line (- - -) between groups indicates the participants were not randomly assigned to conditions. Review Appendix A for more examples of “quasi-experimental” research designs (see also the example of a diagram in Figure 1.2).

Because there is no random assignment in quasi-experimental research, there may be confounding variables influencing the outcome not fully attributed to the treatment (i.e., causal inferences drawn from quasi-experiments must be made with extreme caution). The pretest measure in quasi-experimental research allows the researcher to evaluate the lack of group equivalency and selection bias, thus altering the statistical analysis between experimental and quasi-experimental research for the exact same design (see Cribbie, Arpin-Cribbie, & Gruman, 2010, for a discussion on tests of equivalence for independent group designs with more than two groups).

Figure 1.2 Double Pretest Design for Quasi-Experimental Research

Group	Assignment	Pretest ₁	Pretest ₂	Treatment	Posttest
1	NR	O ₁	O ₂	X	O ₃

2	NR	O ₁	O ₂	—	O ₃

Time ►

Note: This is an example of a between-subjects approach with a double pretest design. The double pretest allows the researcher to compare the “treatment effects” between O₁ to O₂, and then from O₂ to O₃. A major threat to internal validity with this design is testing, but it controls for selection bias and maturation. The two pretests are not necessary if random assignment is used.

It is not recommended to use posttest-only designs for quasi-experimental research. However, if theoretically or logistically it does not make sense to use a pretest measure, then additional controls should be implemented, such as using historical control groups, proxy pretest variables (see Appendix A), or the matching technique to assign participants to conditions.

The reader is referred to Shadish, Clark, and Steiner (2008) for an in-depth discussion of how to use linear regression and propensity scores to approximate the findings of quasi-experimental research to experimental research. They discuss this in the greater context of the potential weaknesses and strengths of quasi-experimental research in determining causation.

CHAPTER 2

BETWEEN-SUBJECTS APPROACH

The between-subjects approach, also known as a *multiple-group* approach, allows a researcher to compare the effects of two or more groups on single or multiple dependent variables (outcome variables). With a minimum of two groups, the participants in each group will only be exposed to one condition (one level of the independent variable), with no crossover between conditions. An advantage of having multiple groups is that it allows for the (a) random assignment to different conditions (experimental research) and (b) comparison of different treatments. If the design includes two or more dependent variables, it can be referred to as a *multivariate approach*, and when the design includes one dependent variable, it is classified as univariate.

PRETEST AND POSTTEST DESIGNS ◆

A common application to experimental and quasi-experimental research is the pretest and posttest between-subjects approach, also referred to as an *analysis of covariance design* (i.e., the pretest measure is used as the *covariate* in the analyses because the pretest should be highly correlated with the posttest). The 1-factor pretest and posttest control group design is one of the most common between-subjects approaches with many

variations (one factor representing one independent variable and sometimes referred to as a *single-factor randomized-group* design). This basic multiple-group design can include a control group and is designed to have multiple measures between and within groups. Although there is a within-subject component, the emphasis is on the between-subject variance. The advantage of including pretest measures allows for the researcher to test for group equivalency (i.e., homogeneity between groups) and for providing a baseline against which to compare the treatment effects, which is the within-subject component of the design (i.e., the pretest is designated as the covariate in order to assess the variance [distance between each set of data points] between the pretest and posttest measures).

There is no set rule that determines the number of observations that should be made on the dependent variable. For example, in a basic pretest and posttest control group design, an observation is taken once prior to the treatment and once after the treatment. However, based on theoretical considerations, the investigator can take multiple posttest treatment measures by including a time-series component. Depending on the research logistics, groups can be randomly assigned or matched, then randomly assigned to meet the criteria for experimental research, or groups can be nonrandomly assigned to conditions (quasi-experimental research). With quasi-experimental research, the limitations of the study significantly increase as defined by the threats to internal validity discussed earlier.

***k*-Factor Designs**

The between-subjects approach can include more than one treatment (factor) or intervention (i.e., the independent variable) and does not always have to include a control group. We designate this design as the *k*-factor design, with or without a control group. Shadish et al. (2002) refer to this design as an alternative- or multiple-treatment design. We prefer the *k*-factor design as a means to clearly distinguish exactly how many factors are present in the design (i.e., the *k* represents the number of factors [independent variables]). To clarify, the treatments in a 3-factor model ($k = 3$), for example, would be designated as X_A , X_B , and X_C (each letter of the alphabet representing a factor) within the design structure. The within-subjects *k*-factor design is referred to as the *crossover* design and is discussed in more detail later in this book under repeated-measures approaches.

A between-subjects *k*-factor design should be used when a researcher wants to examine the effectiveness of more than one type of treatment and a true control is not feasible. Within educational settings, a control group is sometimes not accessible, or there are times when a university's Institutional Review

Board considers the withholding of treatment from specific populations as unethical. Furthermore, some psychologists and educators believe that using another treatment (intervention) as a comparison group will yield more meaningful results, particularly when the types of interventions being studied have a history of proven success; therefore, a *k*-factor design is the obvious choice. We present a variety of examples of 2-, 3-, and 4-factor pretest and posttest designs, as well as posttest-only designs with and without control groups.

Most common threats to internal validity are related, but not limited, to these designs:

Experimental. Maturation, Testing, Attrition, History, and Instrumentation

Quasi-Experimental. Maturation, Testing, Instrumentation, Attrition, History, and Selection Bias

We refer the reader to the following article and book for full explanations regarding threats to validity, grouping, and research designs:

Shadish, W. R., & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, *60*, 607–629.

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

Diagram 2.1 Pretest and Posttest Control Group Design

Group	Pretest	Treatment	Posttest
1	O ₁	X	O ₂
2	O ₁	—	O ₂

Time ►

Note: In regard to design notations, a dashed line (- -) would separate Groups 1 and 2 in the design structure if the participants were not randomly assigned to conditions, which indicates quasi-experimental research.

Example for Diagram 2.1

Chao, P., Bryan, T., Burstein, K., & Ergul, C. (2006). Family-centered intervention for young children at-risk for language and behavior problems. *Early Childhood Education Journal*, *34*(2), 147–153.

Research Question: Does active parent engagement in selecting and using routine-based activities have a positive effect on children's language and appropriate behavior development?

Procedures: The researchers randomly assigned parents to a control and an intervention group. The control group included parents of 19 children, and the intervention group consisted of parents of 22 children. Children in the control group participated only in the pretesting and posttesting phases of the study. Their parents did not receive training and were not required to attend regular meetings or submit weekly and monthly assessments of their children. Parents of children in the family-centered intervention group were trained to use the Child Behavior and Language Assessment (CBLA). Both groups were pretested and posttested on the Test of Early Language Development—Third Edition (TELD-3) and the Eyberg Child Behavior Inventory (ECBI).

Design: Experimental research using a between-subjects approach with a pre- and posttest control group design

Recommended Parametric Analysis: ANCOVA, MANCOVA, two-way RM-ANOVA, or two-way MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Pretest	Treatment	Posttest
R	1 ($n = 22$)	TELD-3, ECBI	Family-centered intervention	TELD-3, ECBI
R	2 ($n = 19$)	TELD-3, ECBI	—	TELD-3, ECBI
Time ►				

Diagram 2.2 2-Factor Pretest and Posttest Control Group Design

Group	Pretest	Treatment	Posttest
1	O_1	X_A	O_2
2	O_1	X_B	O_2
3	O_1	—	O_2
Time ►			

Example for Diagram 2.2

Kazdin, A. E., Esveldt-Dawson, K., French, N. H., & Unis, A. S. (1987). Problem-solving skills training and relationship therapy in the treatment of antisocial child behavior. *Journal of Counseling and Clinical Psychology*, 55(1), 76–85.

Research Question: What are the effects of cognitive-behavioral problem-solving skills training and nondirective relationship therapy on antisocial child behavior?

Procedures: Children were randomly assigned to one of three conditions: (a) problem-solving skills training, (b) relationship therapy, or (c) a treatment-control group. The children met individually for 20 sessions in the two treatment conditions. Sessions lasted approximately 45 minutes and were administered two to three times per week. Treatments were completed while the children were in the hospital. After completion of the sessions, the children were discharged. A treatment-control group was used to partially control for therapist contact and attendance at special sessions outside of the usual ward routine. Children were assessed before and after the intervention with the Child Behavior Checklist (CBCL) and the School Behavior Checklist (SBCL).

Design: Experimental research using a between-subjects approach with a 2-factor pretest and posttest control group design

Recommended Parametric Analysis: ANCOVA, MANCOVA, two-way RM-ANOVA, or two-way MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Pretest	Treatment	Posttest
R	1 (n = 20)	CBCL, SBCL	Problem solving	CBCL, SBCL
R	2 (n = 19)	CBCL, SBCL	Relationship therapy	CBCL, SBCL
R	3 (n = 17)	CBCL, SBCL	—	CBCL, SBCL
Time ►				

Diagram 2.3 2-Factor Pretest and Posttest Design

Group	Pretest	Treatment	Posttest
1	O ₁	X _A	O ₂
2	O ₁	X _B	O ₂

Time ►

Example for Diagram 2.3

Comaskey, E. M., Savage, R. S., & Abrami, P. (2009). A randomized efficacy study of web-based synthetic and analytic programmes among disadvantaged urban kindergarten children. *Journal of Research in Reading*, 32(1), 92–108.

Research Question: Does the ABRACADABRA literacy program produce different effects for synthetic and analytic phonics interventions on phonological, word, and nonword measures?

Procedures: Children were randomly assigned to either the synthetic or analytic phonics intervention group using a manual random-allocation process (allocation cards pulled blind from a hat). This resulted in 27 participants in the analytic phonics group and 26 participants in the synthetic phonics group. Students were engaged in other learning centers and would rotate into the “ABRA” center during the designated time periods. Children would engage in ABRA activities around a single computer, supported by a facilitator. Interventions for the synthetic and analytic phonics groups followed the same lesson structure, beginning with an Animated Alphabet followed by a “core activity.” Three major measures were used in the pre-testing and posttesting sessions, which were the Peabody Picture Vocabulary Test (PPVT) vocabulary scale, Letter-Sound Knowledge (L-SK), and the Wide Range Achievement Test (WRAT).

Design: Experimental research using a between-subjects approach with a 2-factor pretest and posttest design

Recommended Parametric Analysis: ANCOVA, MANCOVA, two-way RM-ANOVA, or two-way MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Pretest	Treatment	Posttest
R	1 ($n = 27$)	PPVT, L-SK, WRAT	Synthetic Phonics	PPVT, L-SK, WRAT
R	2 ($n = 26$)	PPVT, L-SK, WRAT	Analytic Phonics	PPVT, L-SK, WRAT
Time ►				

Diagram 2.4 3-Factor Pretest and Posttest Design

Group	Pretest	Treatment	Posttest
1	O_1	X_A	O_2
2	O_1	X_B	O_2
3	O_1	X_C	O_2
Time ►			

Example for Diagram 2.4

Lee, Y., Park, S., Kim, M., Son, C., & Lee, M. (2005). The effects of visual illustrations on learners' achievement and interest in PDA- (personal digital assistant) based learning. *Journal of Educational Computing Research*, 33(2), 173–187.

Research Question: What are the effects of three types of visual illustrations on learners' achievements, interests, and time spent reading content-specific materials?

Procedures: Participants were randomly assigned to one of three treatment groups: the cognitive interest illustration group, the emotional interest illustration group, and the text-only group. The instructional material containing cognitive interest illustrations consisted of a PDA-based presentation on the topic of the life cycle of the hurricane. Each participant was given the material corresponding to his or her treatment group. A multiple-choice test was used to measure pretest and posttest interest on the topic. Participants then received posttest assessments on achievement and time spent reading the materials.

Design: Experimental research using a between-subjects approach with a 3-factor pretest and posttest design

Recommended Parametric Analysis: ANCOVA, MANCOVA, two-way RM-ANOVA, or two-way MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Diagram 2.5 4-Factor Pretest and Posttest Design

Group	Pretest	Treatment	Posttest
1	O ₁	X _A	O ₂
2	O ₁	X _B	O ₂
3	O ₁	X _C	O ₂
4	O ₁	X _D	O ₂

Time ►

Example for Diagram 2.5

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, 40(1), 281–310.

Research Question: What are the effects of cooperative learning strategies with and without metacognitive training and individualized learning strategies with and without metacognitive training on mathematical reasoning and metacognitive knowledge?

Procedures: Four schools ($N = 384$) were randomly assigned from a pool of 15 schools to one of four treatment conditions: (a) cooperative learning with metacognitive training (COOP+META), (b) cooperative learning (COOP), (c) individualized learning with metacognitive training (IND+META), and (d) individualized learning (IND). All groups received math instruction five times per week with the difference between the groups being the instructional method. The COOP+META condition studied in small heterogeneous

Assignment	Group	Pretest	Treatment	Posttest
R	1 ($n = 15$)	Interest	Cognitive interest	Interest, achievement, time on reading
R	2 ($n = 15$)	Interest	Emotional interest	Interest, achievement, time on reading
R	3 ($n = 15$)	Interest	Text-only	Interest, achievement, time on reading

Time ►

groups using metacognitive strategies. The IND+META condition was the same as the first condition, but students studied individually as opposed to groups. The COOP condition studied in heterogeneous groups but did not use metacognitive training, while the IND group studied individually with no metacognitive training. All groups were administered a graph interpretation test, graph construction test, and a metacognitive questionnaire before and after the completion of the study.

Design: Experimental research using a between-subjects approach with a 4-factor pretest and posttest design

Recommended Parametric Analysis: ANCOVA, MANCOVA, two-way RM-ANOVA, or two-way MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

<i>Assignment</i>	<i>Group</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
R	1 ($n = 105$)	Graph interpretation and construction, metacognitive questionnaire	COOP+META	Graph interpretation and construction, metacognitive questionnaire
R	2 ($n = 95$)	Graph interpretation and construction, metacognitive questionnaire	IND+META	Graph interpretation and construction, metacognitive questionnaire
R	3 ($n = 91$)	Graph interpretation and construction, metacognitive questionnaire	COOP	Graph interpretation and construction, metacognitive questionnaire
R	4 ($n = 93$)	Graph interpretation and construction, metacognitive questionnaire	IND	Graph interpretation and construction, metacognitive questionnaire
Time ►				

POSTTEST DESIGNS ♦

Another approach to experimental research is the between-subjects (or multiple-group) posttest design. The two-group posttest control group design is one of the more common approaches within this structure. By removing

the pretest observation, the within-subject component is eliminated. The main idea behind collecting pretest measures is to ensure group equivalency and control for selection bias, but it also allows for the examination of the differences between baseline (pretest) and posttest measures following the treatment. With posttest-only designs, if random assignment is used, then group equivalency is “secured.” Because only one posttest is observed per condition, this design is not the most rigorous (due to the lack of comparative observations); however, in terms of internal validity, the two-group posttest control group design is considered one of the strongest (see Diagram 2.7).

Research in education does not always allow for conditions suitable for random assignment; therefore, posttest-only designs are not recommended, but they are sometimes the only viable option. Therefore, if random assignment is not used, then a cohort matching technique (i.e., homogeneous groups are assigned to conditions, such as participants from the same class) should be used to assign participants to conditions. The design presented in Diagram 2.6 is a strong alternative for researchers within the field of education who typically cannot use random assignment but have access to groups considered as cohorts. The design includes a between-subjects component and combines a historical control group with a one-group posttest-only design. For example, the researcher can match a group by grade level (i.e., cohort) and then assess the effects of a treatment by contrasting the differences between O_1 of the control and O_1 of the treatment group. An example of the use of this design might include accessing scores on a standardized achievement test (Group 1: O_1) of last year’s seniors with the current senior class’s scores (Group 2: O_1), but only after they received an educational specific intervention (X). Group 1 is designated as a *historical* cohort (i.e., homogeneous) control group.

Diagram 2.6

Example of a Posttest-Only and a Historical Control Group Design

Group	Assignment	Test	Treatment	Posttest
1	NR	O_1	—	—
2	NR	—	X	O_1

Time ►

Note: This design would be designated as quasi-experimental research, hence the NR designation in the chart. History would be the biggest threat to internal validity. An independent-samples t test is the appropriate analysis for this design.

Most common threats to internal validity are related, but not limited, to these designs:

Experimental. Generally, all threats to internal validity are adequately controlled for.

Quasi-Experimental. History, Maturation, Statistical Regression, and Selection Bias

Diagram 2.7 Posttest Control Group Design

Group	Treatment	Posttest
1	X	O ₁
2	—	O ₁

Time ►

Example for Diagram 2.7

Dennis, J. K. (2003). Problem-based learning in online vs. face-to-face environments. *Education for Health, 16*(2), 198–209.

Research Question: What are the differences between online and traditional problem-based learning programs?

Procedures: Students were randomly assigned to one of two conditions: computer-mediated or control (i.e., traditional face-to-face). Students generated learning issues during the initial modules. The students then researched their learning issues and returned with their findings for the second tutorials. Students in the computer-mediated groups interacted with the resource person via e-mail, chat room, or bulletin board only. Following the training, posttest data was collected on learning outcomes (measured by a course examination), time on task (the self-reported time spent in and out of class on learning activities), and generation of learning issues.

Design: Experimental research using a between-subjects approach with a posttest control group design

Recommended Parametric Analysis: Independent-samples *t* test, ANOVA, or MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

<i>Assignment</i>	<i>Group</i>	<i>Treatment</i>	<i>Posttest</i>
R	1 ($n = 17$)	Computer-mediated, problem-based learning	Learning outcomes, time on task, learning issues
R	2 ($n = 17$)	—	Learning outcomes, time on task, learning issues
Time ►			

Diagram 2.8 2-Factor Posttest Control Group Design

<i>Group</i>	<i>Treatment</i>	<i>Posttest</i>
1	X_A	O_1
2	X_B	O_1
3	—	O_1
Time ►		

Example for Diagram 2.8

Green, M., & Cifuentes, L. (2008). An exploration of online environments supporting follow-up to face-to-face professional development. *Journal of Technology and Teacher Education*, 16(3), 283–306.

Research Question: To what extent do participants' attitudes and course completion rates of online teacher professional development programs differ based on the type of training program?

Procedures: Prior to the assignment to each condition, participants were blocked into strata based on service and socioeconomic status (SES). Participants were then randomly assigned (stratified random assignment) to one of three conditions: (a) follow-up with peer interaction, (b) follow-up without peer interaction, and (c) control condition (i.e., traditional training). Each condition included training using the WebCT platform. The two treatment conditions were developed based on the Texas Assessment of Knowledge and Skills standards. The peer interaction group interacted on a weekly basis based on the content from relevant journal articles, websites, and Microsoft PowerPoint, whereas the other group did not interact. All three conditions were administered a posttest assessment on the

attitude of the training modules; completion rates of the workshops were also recorded.

Design: Experimental research using a between-subjects approach with a 2-factor posttest control group design

Recommended Parametric Analysis: ANOVA or MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Treatment	Posttest
R _S	1 (n = 94)	Follow-up with peer interaction	Attitude survey, course completion
R _S	2 (n = 98)	Follow-up without peer interaction	Attitude survey, course completion
R _S	3 (n = 88)	—	Attitude survey, course completion
Time ►			

Diagram 2.9 2-Factor Posttest Design

Group	Treatment	Posttest
1	X _A	O ₁
2	X _B	O ₁
Time ►		

Example for Diagram 2.9

Villalta-Gil, V., Roca, M., Gonzales, N., Domènec, E., Cuca, Escanilla, A., . . . Haro, J. M. (2009). Dog-assisted therapy in the treatment of chronic schizophrenia inpatients. *Antrozoòs*, 22(2), 149–159.

Research Question: Does dog-assisted therapy improve the perceptions of social competence, quality of life, negative and positive symptoms, and satisfaction of schizophrenia inpatients?

Procedures: Patients diagnosed with schizophrenia living in long-term care units were randomly assigned to one of two treatment conditions: Integrated Psychological Treatment (n = 9) and Integrated Psychological Treatment

with a therapy dog ($n = 12$). The treatment intervention included 25 sessions of 45 minutes each following the Integrated Psychological Treatment guidelines. A certified female Labrador therapy dog was used to assist the psychologist with the dog therapy condition. Posttests were administered following the intervention and included the Positive and Negative Symptoms scale, the Living Skills scale, the Brief World Health Organization Quality of Life Assessment, and a Satisfaction With Treatment scale.

Design: Experimental research using a between-subjects approach with a 2-factor posttest design

Recommended Parametric Analysis: independent-samples t test, ANOVA, or MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Treatment	Posttest
R	1 ($n = 12$)	Integrated Psychological Treatment with therapy dog	Positive and Negative Symptoms scale, Living Skills Profile, Quality of Life, Satisfaction
R	2 ($n = 9$)	Integrated Psychological Treatment	Positive and Negative Symptoms scale, Living Skills Profile, Quality of Life, Satisfaction
Time ►			

Diagram 2.10 3-Factor Posttest Design

Group	Treatment	Posttest
1	X_A	O_1
2	X_B	O_1
3	X_C	O_1
Time ►		

Example for Diagram 2.10

Stewart, A. L., King, A. C., & Haskell, W. L. (1993). Endurance exercise and health-related quality of life in 50–65 year-old adults. *The Gerontologist*, 33(6), 782–789.

Research Question: What are the differences between group-based high-intensity training, home-based high-intensity training, and home-based low-intensity training on the VO₂ max, body mass index (BMI), and quality of life indicators of older adults?

Procedures: One hundred and ninety-four participants were randomly assigned to one of three conditions: (a) group-based high-intensity training, (b) home-based high-intensity training, and (c) home-based low-intensity training. The group-based sessions were conducted at a local community college. The home-based training included a first-time face-to-face session and subsequent weekly phone calls for the first 4 weeks, biweekly for the following 4 weeks, and then once a month for the remainder of the 12-month program. Following the yearlong program, participants' VO₂ max, BMI, and perceptions of quality of life (QOL) were assessed.

Design: Experimental research using a between-subjects approach with a 3-factor posttest design

Recommended Parametric Analysis: ANOVA or MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Treatment	Posttest
R	1 (<i>n</i> = 64)	Group-based high-intensity training	VO ₂ , BMI, QOL
R	2 (<i>n</i> = 64)	Home-based high-intensity training	VO ₂ , BMI, QOL
R	3 (<i>n</i> = 64)	Home-based low-intensity training	VO ₂ , BMI, QOL
Time ►			

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. Control is an important element in any type of research. Considering experimental research, come up with a hypothetical research scenario and apply each of the five types of control to the scenario. Use specific examples to illustrate your point.

2. Discuss the three major conditions that must be met in order to establish cause and effect. Next, choose one of these conditions and come up with a scenario that would not allow a researcher to meet this condition. Why would this affect the overall validity of the findings?
3. What makes a posttest-only design “stronger” in terms of internal validity than a design with a pretest? Is it more appropriate to include random assignment for these designs? Why or why not?

Exercise

Develop a hypothetical research scenario that would necessitate the use of a **Pretest and Posttest Control Group Design**. The research will be considered nonexperimental.

1. Identify the research scenario, including the relevant independent and dependent variables.
2. Develop the appropriate primary research question to be associated with this design.
3. Discuss the sampling strategy and technique used to access the appropriate sample.
4. Identify the assignment technique to be used. Discuss whether it will be experimental or quasi-experimental research.
5. Identify what type of comparison group will be used opposite of the treatment group. Why would this type of comparison group be used?
6. In accordance with the assignment technique and comparison group, discuss the various control techniques that will be used with this specific design.
7. Discuss the major threats to validity associated with this design and type of research (experimental or quasi-experimental). How will these threats be addressed, based on the discussion of the control techniques in the previous question?
8. Briefly discuss any limitations associated with this research scenario and the specific design.

CHAPTER 3

REGRESSION- DISCONTINUITY APPROACH

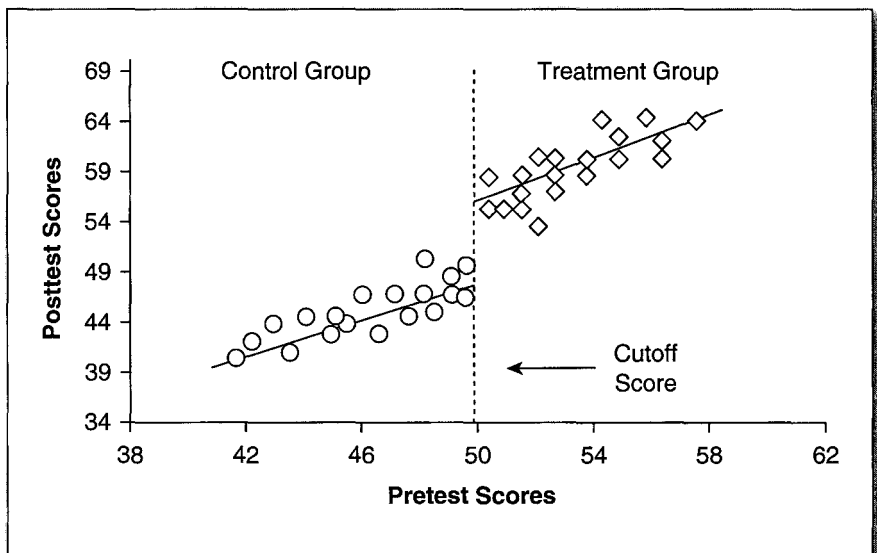
The regression-discontinuity (RD) approach is often referred to as an *RD design*. RD approaches maintain the same design structure as any basic between-subjects pretest and posttest design. The major differences for the RD approach are (a) the method by which research participants are assigned to conditions and (b) the statistical analyses used to test the effects. Specifically, the researcher applies the RD approach as a means of assigning participants to conditions within the design structure by using a cutoff score (criterion) on a predetermined quantitative measure (usually the dependent variable, but not always). Theoretical and logistical considerations are used to determine the cutoff criterion. The cutoff criterion is considered an advantage over typical random or nonrandom assignment approaches as a means to target “needy” participants and assign them to the actual program or treatment condition.

The most basic design used in RD approaches is the two-group pretest-posttest control group design. However, most designs designated as between-subject approaches can use an RD approach as a method of assignment to conditions and subsequent regression analysis. RD approaches can also be applied using data from extant databases (e.g., Luytena, Tymms, & Jones, 2009) as a means to infer causality without designing a true

randomized experiment (see also Lesik, 2006, 2008). As seen in Figure 3.1, the cutoff criterion was 50 (based on a composite rating of 38 to 62). Those who scored below 50 were assigned to the control group, and those who scored above were assigned to the treatment group. As the figure shows, once the posttest scores were collected, a regression line was applied to the model to analyze the pre–post score relationship (i.e., a treatment effect is determined by assessing the degree of change in the regression line in observed and predicted pre–post scores for those who received treatment compared to those who did not).

Some researchers argue that the RD approach does not compromise internal validity to the extent the findings would not be robust to any violations of assumptions (statistically speaking). Typically, an RD approach requires much larger samples as a means to achieve acceptable levels of power (see statistical conclusion validity). We present two examples of studies that employed RD approaches: one that implemented an intervention, and one that used observational data. See Shadish, Cook, and Campbell (2002) for an in-depth discussion of issues related to internal validity for RD approaches, as well as methods for classifying RD approaches as experimental research, quasi-experimental research, and fuzzy regression discontinuity (i.e., assigning participants to conditions in violation of the designated cutoff score).

Figure 3.1 Sample of a Cutoff Score



Most common threats to internal validity are related, but not limited, to these designs:

Experimental. History, Maturation, and Instrumentation

Quasi-Experimental. History, Maturation, Instrumentation, and Selection Bias

We refer the reader to the following articles and book chapter for full explanations regarding RD approaches:

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142, 615–635.

Trochim, W. (2001). Regression-discontinuity design. In N. J. Smelser, J. D. Wright, & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 19, pp. 12940–12945). North-Holland, Amsterdam: Pergamon.

Trochim, W., & Cappelleri, J. C. (1992). Cutoff assignment strategies for enhancing randomized clinical trials. *Controlled Clinical Trials*, 13, 190–212.

Diagram 3.1

Regression-Discontinuity Pretest–Posttest Control Group Design₁

Pretest	Assignment	Group	Treatment	Posttest
O_A	C	1	X	O_2
O_A	C	2	—	O_2

Time ►

Note: O_A refers to the preassignment measure, and C refers to the cutoff score.

Example for Diagram 3.1

Bryant, D. P., Bryant, B. R., Gersten, R., Scammacca, N., & Chavez, M. M. (2008). Mathematic intervention for first- and second-grade students with mathematics difficulties: The effects of tier 2 intervention delivered at booster lessons. *Remedial and Special Education*, 29(1), 20–31.

Research Question: What are the effects of a Tier 2 mathematical intervention on mathematical achievement of first- and second-grade students?

Procedures: The effects of the intervention were determined by including a total of 126 first graders and 140 second graders from a primary-level elementary school. The students were then assessed using the Texas Early Mathematics Inventory–Progress Monitoring (TEMI-PM). Based on the initial results, students who scored at or below the 25th percentile (standard score of 90 or below) were assigned as Tier 2 and subsequently were assigned to the treatment condition. Students who scored above 90 were assigned to the control condition. The result of the cutoff criterion was 26 first graders and 25 second graders who qualified for Tier 2, which included the intervention. The intervention was conceptualized as a booster or supplement to their regular course instruction. This included being exposed to 18 weeks of tutoring sessions. The intervention was grounded in the Texas Essential Knowledge and Skills (TEKS) standards. After the completion of the intervention, all students were administered the TEMI-PM.

Design: Experimental research using an RD approach with a pretest–posttest control group design

Recommended Parametric Analysis: See Imbens and Lemieux (2008) for details.

<i>Pretest</i>	<i>Assignment</i>	<i>Group</i>	<i>Treatment</i>	<i>Posttest</i>
TEMI-PM	C	1 ($n = 51$)	TEKS	TEMI-PM
TEMI-PM	C	2 ($n = 215$)	—	TEMI-PM
Time ►				

Note: Group = First and second graders.

Diagram 3.2

Regression-Discontinuity Pretest–Posttest Control Group Design₂

<i>Pretest</i>	<i>Assignment</i>	<i>Group</i>	<i>Treatment</i>	<i>Posttest</i>
O_A	C	1	X	O_2
O_A	C	2	—	O_2
Time ►				

Example for Diagram 3.2

Leake, M., & Lesik, S. A. (2007). Do remedial English programs impact first-year success in college? An illustration of the regression-discontinuity design. *International Journal of Research & Method in Education*, 30(1), 89–99.

Research Question: What are the effects of participating in a university remedial English program on first-year GPA?

Procedures: A total of 197 first-time university students' scores from an English placement examination were included in this study. An exogenous cutoff score was determined on the placement exam, and students who scored within a 6-point range on either side of the cutoff score on the English placement examination were included in the analysis. Those who scored below the cutoff score, the treatment group ($n = 94$), were required to take a remedial English program. Those who scored above the cutoff score ($n = 103$) did not have to take any remedial courses. Upon completion of the remedial program assignment, the first-year GPA of both groups was included in the RD analysis.

Design: Quasi-experimental research using an RD approach with a pretest–posttest control group design

Recommended Parametric Analysis: See Imbens and Lemieux (2008) for details.

Pretest	Assignment	Group	Treatment	Posttest
English placement examination	C	1 ($n = 94$)	Remedial English program	GPA
English placement examination	C	2 ($n = 103$)	—	GPA
Time ►				

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. What are the advantages to including a cutoff score as a means of assignment in the regression-discontinuity approach?

2. Discuss a research scenario that would warrant the use of a regression-discontinuity approach.

Exercise

Develop a hypothetical research scenario that would necessitate the use of a **Regression-Discontinuity Posttest-Only Control Group Design**. The research will be considered nonexperimental.

1. Identify the research scenario, including the relevant independent and dependent variables.
2. Develop the appropriate primary research question to be associated with this design.
3. Discuss the sampling strategy and technique used to access the appropriate sample.
4. Identify the variable and what criteria will be used as the cutoff score for group assignment. Discuss whether it will be experimental or quasi-experimental research.
5. In accordance with the assignment technique and comparison group, discuss the various control techniques that will be used with this specific design.
6. Discuss the major threats to validity associated with this design and type of research (experimental or quasi-experimental). How will these threats be addressed, based on the discussion of the control techniques in the previous question?
7. Briefly discuss any limitations associated with this research scenario and the specific design.

CHAPTER 4

WITHIN-SUBJECTS APPROACH

Major challenges when conducting research are often related to (a) access to participants and (b) an inability to randomly assign the participants to conditions. With these limitations in mind, researchers often employ a within-subjects approach. Although the pretest and posttest designs of between-subjects approaches include a within-subject component, the objective is not necessarily to test the within-subject variances as intended with within-subject approaches. The within-subjects approach to research assumes one group (or subject) serves in each of the treatment conditions.

This approach is referred to as *repeated measures* because participants are repeatedly measured across each condition. The advantage to this approach is that it can be used with smaller sample sizes with little or no error variance concerning individual differences between conditions (i.e., the same participants exist in each condition). Some disadvantages to this approach are the threats to internal validity, which are primarily maturation and history, and the biggest issue is sequencing effects (i.e., order and carryover effects). More specifically, performance in one treatment condition affects the performance in a second treatment condition. If possible, it is recommended to randomize the order of the treatments (also known as *counterbalancing*) to control for sequencing effects.

The simplest within-subjects approach is the one-group with a single pretest and posttest measure (quasi-experimental research 1-factor design),

which is presented here. This design can be extended to multiple pretest and posttest measures and is designated as an interrupted time-series (ITS) design and is sometimes called the “time-series” approach. For this guide, we categorize the ITS design under the repeated-measures approach. Traditionally, it was believed that ITS designs should include upward of 100 observations (in regard to statistical power), but many of these designs, when applied, often have anywhere from 10 to 50 observations and are often designated as short ITS designs.

♦ REPEATED-MEASURES APPROACH

The repeated-measures approach is structured so the researcher can collect numerous measures from the participants. Specifically, designs that include repeated measures allow researchers to gather multiple data points over time to study the rate of change as a function of treatment or time. These types of designs typically are more advanced, which require advanced statistical analysis to summarize the data. Most single-case approaches must use repeated-measures approaches. This approach allows for the single unit of analysis to serve as its own control to minimize treatment effects. Designs that employ repeated-measures approaches are also useful in longitudinal studies when examining trends or phenomena over a designated period of time. There are several designs that use the repeated-measures approach.

It is important to clarify that designs within the repeated-measures approach are classified as *experimental* as long as participants are randomly exposed to each condition (i.e., counterbalancing must occur because sequencing effects are the biggest threat to internal validity within this approach). However, there are repeated-measures approaches that are considered nonexperimental research. The ITS design is an example of nonexperimental research and is often referred to as a *longitudinal data structure* because data is collected at varying time points over days, months, or even years. The application of this approach, as with all approaches, is considered along with theoretical tenets and logistical considerations.

Repeated-measures approaches can also include a between-subjects component as seen in the pretest and multiple-posttest design and the switching-replications design (the emphasis is usually on the between- and within-subject variances, which are sometimes not referred to as repeated measures because technically each group is not exposed to each condition). We present one example of the pretest and multiple-posttest design

and two examples of a switching-replication design (one experimental and one quasi-experimental). This design allows the researcher to assess the effects of the treatment on the first group while withholding the treatment to the second group. The second group is designated as a wait-list control group. This design includes only one treatment or factor. We also present a similar design, the crossover design (also known as a *changeover design*), which includes at a minimum two factors, but it can include more (Ryan, 2007; Shadish, Cook, & Campbell, 2002). Some researchers, as seen in the experimental example presented later, refer to a switching-replications design as a crossover design. To be clear, the switching-replications design includes one treatment and a wait-list control group, while the crossover design includes a minimum of two treatments and no control.

Switching-Replications Design: A Primer

The switching-replications design is one of the most effective experimental designs at controlling for threats to internal validity. Perhaps more importantly, it eliminates the need to deny any potentially beneficial intervention to participants due to random assignment (to control group). The design is straightforward: The treatment is replicated (repeated) with each group, with one group receiving the treatment first. In theory, external validity should also be improved through the use of two independent administrations of the same intervention. Treatment environment and condition always vary somewhat over time (outside of a laboratory setup), thus having the treatment replicated at a later time (with the potential of many variations in the treatment application and environment and history) with similar results would demonstrate generalizability.

However, the standard design structure for a switching-replication design should not be chosen if the research can use random assignment to conditions because it is nearly impossible to avoid violating the standard statistical assumptions associated with repeated-measures analysis. Therefore, we propose a variant called the *wait-list continuation design* when random assignment is available for application. The design includes both a within- and between-subjects component (i.e., mixed-subjects approach). A wait-list control group is incorporated and doesn't include the pretest for that condition. In effect, each group serves as both treatment and control at different points in time and allows for statistical analysis without relying on statistical assumptions to fall into place naturally (e.g., multivariate normality and sphericity). We provide a mock statistical analysis of this design in Appendix E.

Figure 4.1 Wait-List Continuation Design (for Random Assignment)

<i>Group</i>	<i>Treatment</i>	<i>Test</i>	<i>Treatment</i>	<i>Test</i>	-	<i>Test</i>
1	X_A	O_1	- <i>Time Delay</i>	O_2	-	-
2	-	O_3	X_A	O_4	- <i>Time Delay</i>	O_5

Time ►

The quasi-experimental application of this design works well in educational environments where programs are repeated at standard intervals throughout the year. The public school system is a good example. Specifically, most school systems work on a semester structure, which allows for one group of students to receive an intervention during the first semester and a second group of students to receive the same intervention during the second semester. When resources (teachers, space, etc.) are limited, this affords an organization the opportunity to provide a program to a greater number of students over time. Since all participants eventually participate, this design is often the most ethical (as well as feasible) of all the quasi-experimental designs.

Figure 4.2 Switching-Replications Design (for Nonrandom Assignment)

<i>Group</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Midtest</i>	<i>Treatment</i>	<i>Posttest</i>
1	O_1	X_A	O_2	-	O_3
2	O_4	-	O_5	X_A	O_6

Time ►

There are a couple of important factors to consider when contemplating a switching-replications design. Specifically, is the intervention something that would theoretically (or desirably) have a lasting impact, or does the researcher want the intervention to maintain its effect, even after the treatment session is complete? If so, a critical part of the analysis is to examine what occurs between Time Point 2 and Time Point 3 of the group that first receives treatment; in this case, the within-subjects analysis would be primary and between-subjects secondary. If, however, the treatment is

something that would not theoretically continue to impact the dependent variable (e.g., the impact of aspirin on a headache), then theoretically all of the data from the control condition (from the two groups) should be collapsed, and all of the data from the treatment condition (from the two groups) should be collapsed and a between-subjects analysis used as the primary analysis with the within-subjects being secondary. In both cases a between- and a within-subjects analysis should be used when conducting a switching-replications design.

Crossover and Latin-Square Designs

Crossover designs are the repeated-measures version of the k -factor design and are used to assess the order of effects of two or more factors (also known as *multiple-treatment counterbalanced designs*). When applying crossover designs, it is important to ensure a “washout period,” or return to baseline, between the adjacent treatment periods as a means to control for sequencing effects (i.e., multiple-treatment interference). These types of designs are ideal for eliminating issues associated with the between-subject variations and when a limited number of test subjects are available. However, if there are enough subjects to assign to groups to each condition, as seen in the 2-factor example presented later, then a between- and within-subject analysis should be used. Alternatively, a 3-factor model ($k = 3$; see Diagram 4.1), with one participant assigned to each condition, would require only a within-subjects analysis. The reader is referred to Hedayat, Stufken, and Yang (2006) for more examples of crossover designs.

We also present two examples of ITS designs. The basic ITS design includes one treatment (or factor) and many consecutive observations on the same outcome variable prior to and after the treatment. The number of pretest and posttest observations is based on theoretical, logistical, and statistical considerations. The first example includes two groups and one factor, and the second example includes one group and one factor. The strength of the ITS design is that it can account for the immediate and/or delayed effects of a treatment. The largest threats to internal validity for ITS designs are history and attrition. See Glass, Wilson, and Gottman (2008) for in-depth coverage of ITS designs and analysis.

Last, similar to the crossover design, we present an example of an $n \times n$ Latin-square design. The Latin-square design is a one-factor model with two nuisance or procedural factors (one for the rows and one for the columns), and it is most commonly applied in engineering, agriculture, and industrial research but rarely in the social sciences. However, there are

Diagram 4.1 A Repeated-Measures Approach 3-Factor Crossover Design

Subject	Treatment	Midtest	Treatment	Midtest	Treatment	Posttest
1	X _A	O ₁	X _B	O ₂	X _C	O ₃
2	X _B	O ₁	X _C	O ₂	X _A	O ₃
3	X _C	O ₁	X _A	O ₂	X _B	O ₃
4	X _C	O ₁	X _B	O ₂	X _A	O ₃
5	X _A	O ₁	X _C	O ₂	X _B	O ₃
6	X _B	O ₁	X _A	O ₂	X _C	O ₃

Time ►

Note: Each participant ($N = 6$) serves in one condition, and the conditions are counterbalanced to control for sequencing effects. This design can be modified in multiple ways, such as adding additional factors, introducing the same factor more than once in each condition, and including more observations.

instances within educational and the social and behavioral sciences in which a Latin-square design may be used. Within the Latin-square structure, each row and each column contain the treatment as a means to counterbalance the order of effects. A basic 3×3 Latin-square design can be applied, for example, if a researcher wishes to examine the effectiveness of three types of emotive imagery techniques (strong, medium, weak) on professional athletes' level of concentration (assuming a total of 75 athletes divided between each condition is adequate with regard to power). As seen in Diagram 4.2, the researcher would use three different settings (office, home, field) and three types of formats (live, recorded, combination) to administer the technique. This one-factor design includes emotive imagery (at three levels) and then two procedural factors (format and setting), each at three levels, hence the 3×3 framework.

This design is best applied in highly controlled conditions, and it is typically used as a means to protect against the effects of multiple extraneous variables. Some researchers use some form of a crossover design in their research and set up the rows and columns in the form of a Latin square. However, without the inclusion of two procedural (blocking) factors, the examination does not take full advantage of the design characteristics for which the Latin-square design was originally created. The major assumption of this design is that there is no interaction (or very

Diagram 4.2 Example of a 3×3 Latin-Square Design

Group	Setting	Format		
		Live	Recorded	Combo
1 ($n = 25$)	Office	A	B	C
2 ($n = 25$)	Home	B	C	A
3 ($n = 25$)	Field	C	A	B

Note: A = Strong emotive imagery; B = Medium emotive imagery; C = Weak emotive imagery. Concentration would be assessed within each session. A general linear model one-way ANOVA is the appropriate analysis for this design.

minimal) between rows and columns. The problem associated with this design when applied to the social sciences is the small number of observations for each combination of factor levels; in addition, when including humans as test subjects, the carryover effects are usually problematic (i.e., the efficiency of this design is application and data dependent). We refer the reader to McNemar (2007) for a discussion regarding the issues associated with the use of Latin-square designs in the social sciences. Based on the example scenario presented earlier, not every combination of imagery-format-setting is conducted. Theoretically, a $3 \times 3 \times 3$ factorial design may be a more suitable design for this application. Nonetheless, if the researcher wishes to include a third procedural factor (type of athlete, for example), then a Graeco-Latin square design should be used (see Box, 2009, for more on Graeco-Latin square designs). We also refer the reader to Reese (1997) for more information regarding the applications and analyses of Latin-square designs. We later present an example of a 6×6 Latin-square design.

Most common threats to internal validity are related, but not limited, to these designs:

Experimental. History, Maturation, Testing, Instrumentation, Attrition, and Sequencing Effects

Quasi-Experimental. History, Maturation, Testing, Instrumentation, Statistical Regression, Selection Bias, Attrition, and Sequencing Effects

We refer the reader to the following books for full explanations regarding research for repeated-measures approaches:

Hinkelman, K. (2012). *Design and analysis of experiments, Vol. 3: Special designs and applications*. Hoboken, NJ: Wiley.

Spector, P. E. (1981). *Research designs*. Beverly Hills, CA: Sage.

Diagram 4.3 Pretest and Posttest Design (One-Group)

Group	Pretest	Treatment	Posttest
1	O ₁	X	O ₂
Time ►			

Example for Diagram 4.3

Wyatt, T. H., & Hauenstein, E. J. (2008). Pilot testing Okay With Asthma™: An online asthma intervention for school-age children. *The Journal of School Nursing, 24*(3), 145–150.

Research Question: What are the effects of the Okay With Asthma intervention on asthma knowledge and attitude in school-age children?

Procedures: Thirty-five school-age children were recruited to participate in the study. During the first session, children completed the pretests and Okay With Asthma program under the supervision of the investigator. The investigator conducted a debriefing with each child after he or she viewed the multimedia program. Following the end of the program, the children completed a round of posttest measures. The pretest and posttest measures included the Asthma Information Quiz and the Child Attitude Toward Illness Scale.

Design: Quasi-experimental research using a within-subjects approach and a one-group pretest and posttest design

Recommended Parametric Analysis: Descriptive statistics; dependent-samples *t* test, or paired-samples *t* test (appropriate effect-size calculations should be included)

Assignment	Group	Pretest	Treatment	Posttest
NR	1 ($n = 35$)	Asthma Information Quiz, Child Attitude Toward Illness Scale	Okay With Asthma™	Asthma Information Quiz, Child Attitude Toward Illness Scale
Time ►				

Diagram 4.4 Pretest and Multiple-Posttest Design

Group	Pretest	Treatment	Posttest ₁	Posttest ₂
1	O ₁	X	O ₂	O ₃
2	O ₁	—	O ₂	O ₃
Time ►				

Note: Any number of posttests and factors can be included, based on theoretical and logistical considerations.

Example for Diagram 4.4

Acee, T. W., & Weinstein, C. E. (2010). Effects of value-reappraisal intervention on statistics students' motivation and performance. *Journal of Experimental Education*, 78(4), 487–512.

Research Question: What are the effects of a value-reappraisal intervention on task value, endogenous instrumentality, self-efficacy, and exam performance?

Procedures: Participants were stratified based on instructor type, gender, and year in school and then randomly assigned to each condition: the value-reappraisal (VR) group ($n = 41$) and the control group ($n = 41$). Both groups were assessed prior to the intervention on task value (TV), endogenous instrumentality (EI), self-efficacy (SE), and exam performance (EP). For the next 3 weeks, the students in the treatment group were exposed to the VR intervention, which was designed to help them reappraise their values of the statistics course. Immediately following the intervention, both groups received the TV, EI, SE, and EP assessments. Following a 2-week delay, another round of the assessments was given.

Design: Experimental research using a repeated-measures approach with a pretest and double-posttest design

Recommended Parametric Analysis: One-way RM-ANOVA or one-way MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Pretest	Treatment	Posttest	Posttest (2-Week Delay)
R _s	1 (n = 41)	TV, EI, SE, EP	Value-Reappraisal	TV, EI, SE, EP	TV, EI, SE, EP
R _s	2 (n = 41)	TV, EI, SE, EP	—	TV, EI, SE, EP	TV, EI, SE, EP
Time ►					

Diagram 4.5 Switching-Replications Design (Experimental)

Group	Pretest	Treatment	Midtest	Treatment	Posttest
1	O ₁	X _A	O ₂	—	O ₃
2	O ₁	—	O ₂	X _A	O ₃
Time ►					

Example for Diagram 4.5

Cernin, P. A., & Lichtenberg, P. A. (2009). Behavioral treatment for depressed mood: A pleasant events intervention for seniors residing in assisted living. *Clinical Gerontologist*, 32, 324–331.

Research Question: What are the effects of pleasant-events-focused treatment on factors of mood and depression in older, frail adults?

Procedures: Participants were randomly assigned to either the immediate treatment condition ($n = 8$) or wait-list treatment condition ($n = 7$). Both groups received the same events-focused treatment. Data were collected by the project coordinator for all participants at baseline, 3 months, and 6 months. Participants enrolled in the immediate treatment condition received the intervention after baseline data collection in the first 3 months of the study. Participants enrolled in the wait-list treatment condition received the intervention between 3 and 6 months. Treatment consisted of pleasant-event activities mutually agreed upon with residents and was delivered in 30-minute sessions with a target goal of three sessions per week.

Design: Experimental research using a repeated-measures approach with a switching-replications design

Recommended Parametric Analysis: One-way RM-ANOVA or a combination of independent- and dependent-samples *t* tests (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Pretest (Baseline)	Treatment	Midtest (3 Months)	Treatment	Posttest (6 Months)
R	1 (<i>n</i> = 8)	Mood, depression	Pleasant-events-focused treatment	Mood, depression	—	Mood, depression
R	2 (wait-list; <i>n</i> = 7)	Mood, depression	—	Mood, depression	Pleasant-events-focused treatment	Mood, depression
Time ►						

Diagram 4.6 Switching-Replications Design (Quasi-Experimental)

Group	Pretest	Treatment	Midtest	Treatment	Posttest
1	O ₁	X _A	O ₂	—	O ₃
2	O ₁	—	O ₂	X _A	O ₃
Time ►					

Example for Diagram 4.6

Basadur, M., Graen, G. B., & Scandura, T. A. (1986). Training effects on attitudes toward divergent thinking among manufacturing engineers. *Journal of Applied Psychology*, 71(4), 612–617.

Research Question: What are the effects of a training process of problem solving based on divergent thinking on engineers' attitudes?

Procedures: Initially, two groups of manufacturing engineers were identified as participants for the study. Group 1 (*n* = 65) was assigned to the

first treatment condition (creative problem-solving program) and Group 2 ($n = 47$) was assigned to the wait-list control condition. Both groups were administered scales that measured the preferences for ideation in problem solving and a second scale that measured the premature convergence in problem solving. After 5 weeks, the measures were administered to both groups, and then the wait-list control went through the creative problem-solving program. Following the next 5 weeks, participants were administered the final posttest measures.

Note: The authors refer to this design as a crossover design; however, to qualify as a crossover design, it would require a minimum of two factors or treatments. Therefore, this is a switching-replications design (Ryan, 2007; Shadish et al., 2002).

Design: Quasi-experimental research using a repeated-measures approach with a switching-replications design

Recommended Parametric Analysis: Combination of independent- and dependent-samples t tests (The RM-ANOVA is not recommended because of the obvious violations of the specific statistical assumptions; appropriate descriptive statistics and effect-size calculations should be included.)

Assignment	Group	Pretest	Treatment	Midtest	Treatment	Posttest
NR	1 ($n = 65$)	Preference for ideation, premature evaluations	Creative problem-solving program	Preference for ideation, premature evaluations	—	Preference for ideation, premature evaluations
NR	2 (wait-list; $n = 47$)	Preference for ideation, premature evaluations	—	Preference for ideation, premature evaluations	Creative problem-solving program	Preference for ideation, premature evaluations
Time ►						

Diagram 4.7 Crossover Design (2-Factor)

Group	Pretest	Treatment	Midtest	Treatment	Posttest
1	O_1	X_A	O_2	X_B	O_3
2	O_1	X_B	O_2	X_A	O_3
Time ►					

Example for Diagram 4.7

Burgess, G., Grogan, S., & Burwitz, L. (2006). Effects of a 6-week aerobic dance intervention on body image and physical self-perceptions in adolescent girls. *Body Image*, 3, 57–66.

Research Question: What are the effects of an aerobic and physical education program on perceptions of body image in adolescent girls?

Procedures: A total of 50 school-age girls ($M_{age} = 13.5$) were randomly assigned to one of two conditions. The first condition was an aerobic dance program, and the second condition was physical education. Each condition lasted approximately six weeks. Following the completion of the first program, the participants changed over (crossed over) to participate in the other treatment program. All participants were administered pretest, midtest, and posttest the Body Attitudes Questionnaire (BAQ), the Children and Youth Physical Self-Perceptions Profile (CY-PSPP), and the Leisure Time Physical Activity Questionnaire (LTPAQ).

Design: Experimental research using a repeated-measures approach with a 2-factor crossover design

Assignment	Group	Pretest	Treatment	Midtest	Treatment	Posttest
R	1 (n = 25)	BAQ, CY-PSPP, LTPAQ	Aerobic dance	BAQ, CY-PSPP, LTPAQ	Physical education	BAQ, CY-PSPP, LTPAQ
R	2 (n = 25)	BAQ, CY-PSPP, LTPAQ	Physical education	BAQ, CY-PSPP, LTPAQ	Aerobic dance	BAQ, CY-PSPP, LTPAQ
Time ►						

Diagram 4.8 Interrupted Time-Series Design (One-Group)

Group	Pretests	Treatment	Posttests
1	$O_1 \dots O_8$	X	$O_7 \dots O_{10}$
Time ►			

Note: Any number of pretest and posttest observations can be taken in an ITS design.

Recommended Parametric Analysis: One-way RM-ANOVA or one-way MANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Example for Diagram 4.8

Rimondini, M., Del Piccolo, L., Goss, C., Mazzi, M., Paccaloni, M., & Zimmerman, C. (2010). The evaluation of training in patient-centered interviewing skills for psychiatric residents. *Psychological Medicine, 40*, 467–476.

Research Question: What is the impact of patient-centered interview training on patient-centered interviewing skills?

Procedures: A single group of psychiatric residents ($n = 10$) interviewed a total of 12 different patients. Each resident interviewed 8 patients and was objectively tested by trained raters who used the Verona Psychiatric Interview Classification System (VR-PICS). The residents then went through a patient-centered skills training using the standards of the Verona Communication Skills Training in Psychiatry. The residents were then interviewed and were rated on a total of four more time points using the VR-PICS. The major categories tested were gathering information, handling emotion, and handling doctor-centered issues, as well as an overall performance index.

Design: Quasi-experimental research using a one-group interrupted time-series design

Assignment	Group	Pretests 1–8	Treatment	Posttests 7–10
NR	1	VR-PICS	Verona Skills Training	VR-PICS
Time ►				

Diagram 4.9 Interrupted Time-Series Design (Two-Group)

Group	Pretests	Treatment	Posttests
1	$O_1 \dots O_6$	X	$O_7 \dots O_{11}$
2	$O_1 \dots O_6$		$O_7 \dots O_{11}$
Time ►			

Recommended Parametric Analysis: Growth curve analysis, autoregressive integrated moving average (ARIMA), or time-series regression (appropriate descriptive statistics and effect-size calculations should be included)

Example for Diagram 4.9

May, H., & Supovitz, J. A. (2006). Capturing the cumulative effects of school reform: An 11-year study of the impacts of America's Choice on student achievement. *Educational Evaluation and Policy Analysis*, 28(3), 231–257.

Research Question: What is the evidence of, and rate of increase in, learning in low-achieving and minority students in America's Choice programs?

Procedures: Eleven years of student performance school data were analyzed: 6 years of data before schools' adoption of the America's Choice program and up to 5 years after (depending on the year a school introduced the reform). During this period, more than 55,000 students in Grades 1 through 8 in 42 elementary and 10 middle schools were tested in reading and mathematics. Multiple achievement tests during the 11-year span covered the study: (a) Stanford Achievement Test (SAT-9), (b) California Achievement Test (CAT-5), (c) Degrees of Reading Power test (DRP), (d) New York State assessments (NYS), (e) New York Pupil Evaluation Program tests (PEP), and (f) New York Preliminary Competency Test (PCT).

Design: Quasi-experimental research using a two-group interrupted time-series design

Recommended Parametric Analysis: Growth curve analysis, autoregressive integrated moving average (ARIMA), or time-series regression (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Years 1–6	Treatment	Years 7–11
NR	1 (Reading: $n = 56,693$)	SAT-9, CAT-5, DRP, NYS, PEP, PCT	America's Choice	SAT-9, CAT-5, DRP, NYS, PEP, PCT
NR	2 (Math: $n = 55,932$)	SAT-9, CAT-5, DRP, NYS, PEP, PCT		SAT-9, CAT-5, DRP, NYS, PEP, PCT
Time ►				

Diagram 4.10 6 × 6 Latin-Square Design

<i>Group</i>	<i>Conditions</i>					
1	A	B	C	D	E	F
2	B	C	F	A	D	E
3	C	F	B	E	A	D
4	D	E	A	B	F	C
5	E	A	D	F	C	B
6	F	D	E	C	B	A

Time ►

Note: There are many possible ordering variations of treatments in the Latin-square design, and the design presented here is only one example of a type of ordering.

Example for Diagram 4.10

Meijer, J., & Oostdam, R. (2007). Test anxiety and intelligence testing: A closer examination of the stage-fright hypothesis and the influence of stressful instruction. *Anxiety, Stress, and Coping*, 20(1), 77–91.

Research Question: What are the effects of text anxiety and style of instruction (stressful versus reassuring) on intelligence?

Procedures: A series of intelligence tests were administered to students over a 1-year time span. There were 135 students across six classes, and each class varied in size from 17 to 26 children. The age range was between 10 and 13 years old. The intelligence tests were taken from a version of the Netherlands Differentiation Test and included a vocabulary test, picture recall, dice and dominoes, verbal analogies, paired associates, and figure series. The administration of each test was counterbalanced for each class according to the Latin-square design. The first three tests for each class were administered under stressful conditions and the last three were administered under reassuring conditions. The Revised Worry-Emotionality Scale (RW-ES) was used to determine the level of perceived stress prior to each series of tests.

Design: Experimental research using a 6 x 6 Latin-square design

Recommended Parametric Analysis: General linear models one-way ANOVA (appropriate descriptive statistics and effect-size calculations should be included)

Class (N = 135)	Stressful Instruction	Intelligence Tests			Reassuring Instruction	Intelligence Tests		
		Voc	Pic	Dice		Ana	Pair	Fig
1	RW-ES	Voc	Pic	Dice	RW-ES	Ana	Pair	Fig
2	RW-ES	Pic	Dice	Ana	RW-ES	Pair	Fig	Voc
3	RW-ES	Dice	Ana	Pair	RW-ES	Fig	Voc	Pic
4	RW-ES	Ana	Pair	Fig	RW-ES	Voc	Pic	Dice
5	RW-ES	Pair	Fig	Voc	RW-ES	Pic	Dice	Ana
6	RW-ES	Fig	Voc	Pic	RW-ES	Dice	Ana	Pair
Time ►								

Note: Voc = Vocabulary; Pic = Picture recall; Dice = Dice and dominoes; Ana = Verbal analogies; Pair = Paired associates; Fig = Figure series. This design is ordered as a Latin square; however, it does not follow the exact tenets of a Latin-square design (the lack of two blocking factors). Therefore, it can be considered a modified version of a repeated-measures 6-factor crossover design configured as a Latin square.

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. Discuss several reasons why a researcher would choose a within-subjects approach over a between-subjects approach.
2. Discuss three of the major threats to validity related to the repeated-measures approach. What is it about this approach that would make these threats important to consider?
3. Develop a hypothetical research scenario that would necessitate the use of a 3-factor crossover design. Why would this design be the best option to utilize?

Exercise

Develop a hypothetical research scenario that would necessitate the use of a **3-Factor Crossover Design**. The research will be considered nonexperimental.

1. Identify the research scenario, including the relevant three independent variables and dependent variable(s).
2. Develop the appropriate primary research question to be associated with this design.

3. Discuss the sampling strategy and technique used to access the appropriate sample.
4. Identify the assignment technique to be used.
5. In accordance with the assignment technique and comparison group, discuss the various control techniques that will be used with this specific design.
6. Discuss the major threats to validity associated with this design and type of research (experimental). How will these threats be addressed, based on the discussion of the control techniques in the previous question?
7. Briefly discuss any limitations associated with this research scenario and the specific design.

CHAPTER 5

FACTORIAL DESIGNS

An extension of the k -factor design is the factorial design. The simplest factorial design includes, at a minimum, two factors (i.e., independent variables), each with two levels (Kazdin, 2002; Vogt, 2005). Two factors each with two levels is designated as a 2×2 factorial design. Factorial designs are denoted by the form s^k . The s represents the number of levels, and k represents the number of factors (e.g., 2×2 is the same as 2^2). Recall that a factor is another term for the independent variable, or treatment, or intervention.

Many k -factor designs can be transformed into factorial designs (based on theoretical and logistical considerations) by partitioning the factors into at least two levels and by subsequently changing the statistical analysis used to examine the data. For example, a researcher is interested in looking at the effects of a math intervention (1: factor) partitioned into two levels (1–visual math, 2–auditory math) and how it differs by gender (2: factor; 1–males, 2–females) on a math competency exam (i.e., dependent variable). Unlike the k -factor design, factorial designs allow for *all* combinations of the factor levels to be tested on the outcome (i.e., male and female differences for auditory-style teaching compared to male and female differences for visual-style teaching). Thus, factorial designs allow for the examination of both the *interaction* effect (the influence of one independent variable on the other independent variable) and the *main* effects (the influence of each independent variable on the outcome).

We must caution the social and behavioral science researcher not to get overzealous with the application of more complex factorial designs outside of the 2×2 design. A general assumption related to the factorial design is

that there is no interaction between the factors, but this is typically impossible when including humans as test subjects. The factorial design was originally developed for agricultural and engineering research where the variables are static (e.g., amount of fertilizer or blade length) and does not suffer from the typical threats to internal validity that occurs when human participants are the test subjects (e.g., testing, sequencing effects).

The factorial design is not one design; rather, it is considered a family of designs. For example, some research requires that the number of levels for each factor is not the same. The simplest version would be a 2×3 design (i.e., one independent variable has two levels and the other has three). Factorial designs can also include three factors (e.g., $2 \times 2 \times 2$ represents three independent variables, each with two levels). Factorial designs can use within-subjects or between-subjects approaches, and they can include pretest and posttest or posttest-only measures (most contain only posttests). The within-subjects approach to factorial designs is set up so there is one group, and each participant serves in each of the treatment conditions. The between-subjects approach allows the researcher to test multiple groups across conditions without exposing each participant to all treatment conditions. This approach requires larger sample sizes, and random assignment is highly recommended to control for differentiation and selection bias.

Another option to the factorial design is the mixed-subjects approach. A mixed-factorial design includes both a within- and between-subjects approach. For instance, a 2×3 mixed-factorial design would be constructed so the first factor at two levels is tested as within subjects, and the second factor at three levels would be tested as between subjects. To determine the number of groups (also referred to as *cells*), the number of levels for each factor can be multiplied (e.g., $2 \times 2 = 4$ groups; $2 \times 3 = 6$ groups). The strength of this design is that it allows a researcher to examine the individual and combined effects of the variables. There are many types and variations of factorial designs not illustrated in this reference guide.

We provide three examples of a 2-factor between-subjects factorial design (a pretest and posttest design [2×2] and two posttest-only designs [2×2 and 3×2]) and one example of a 2×2 within-subjects factorial design. We also provide two examples of a between-subjects factorial design with three factors ($2 \times 2 \times 2$ and $2 \times 3 \times 2$) and one example of a mixed-factorial design ($2 \times 2 \times 2$).

Factorial designs that include within-subjects components are also affected by the threats to internal validity listed under the repeated-measures approach (e.g., sequencing effects).

Most common threats to internal validity are related, but not limited, to these designs:

Experimental. Maturation, Testing, Diffusion, and Instrumentation

Quasi-Experimental. Maturation, Testing, Instrumentation, Diffusion, and Selection Bias

We refer the reader to the following article and book for full explanations regarding factorial designs:

Dasgupta, T., Pillai, N. S., & Rubin, D. B. (2014). Causal inference from 2^K factorial designs using potential outcomes. *Journal of the Statistical Society: Series B (Statistical Methodology)*, 77(4), 727–753.

Ryan, T. P. (2007). *Modern experimental design*. Hoboken, NJ: Wiley.

Diagram 5.1

2×2 Factorial Pretest and Posttest Design (Between-Subjects)

Group	Pretest	Treatment	Posttest
1	O_1	X_{A1B1}	O_2
2	O_1	X_{A1B2}	O_2
3	O_1	X_{A2B1}	O_2
4	O_1	X_{A2B2}	O_2

Time ►

Example for Diagram 5.1

Stern, S. E., Mullennix, J. W., & Wilson, S. J. (2002). Effects of perceived disability on persuasiveness of computer-synthesized speech. *Journal of Applied Psychology*, 87(2), 411–417.

Research Question (main and interaction effects): What are the effects of perceived disabilities on the persuasiveness of computer-synthesized and normal speech?

Procedures: Participants completed an attitude pretest and were randomly assigned to watch an actor deliver a persuasive speech under one of the following four conditions: (a) disabled using normal speech, (b) nondisabled using normal speech, (c) disabled using computer-synthesized speech, or (d) nondisabled using computer-synthesized speech. Participants then completed an attitude posttest survey. Additionally, the following

posttest measures not included in the factorial analysis were collected: (a) questionnaires assessing attitudes, (b) perceptions of voice characteristics, and (c) the effectiveness of the message.

Design: Experimental research using a between-subjects approach 2×2 factorial pretest and posttest design

Recommended Parametric Analysis: 2-way between-subjects ANCOVA or 2-way MANCOVA (appropriate follow-up analysis plus descriptive statistics and effect-size calculations should be included)

Assignment	Group	Pretest	Factor Ability (A)	Factor Speech Type (B)	Posttest
R	1	Attitudes	Disabled (A_1)	Synthetic speech (B_1)	Attitudes, speech characteristics, effectiveness of message
R	2	Attitudes	Disabled (A_1)	Normal speech (B_2)	Attitudes, speech characteristics, effectiveness of message
R	3	Attitudes	Nondisabled (A_2)	Synthetic speech (B_1)	Attitudes, speech characteristics, effectiveness of message
R	4	Attitudes	Nondisabled (A_2)	Normal speech (B_2)	Attitudes, speech characteristics, effectiveness of message
Time ►					

2×2 ($N = 189$)	<i>Independent Variable</i> Ability (A)	
	<i>Disabled</i> (A_1)	<i>Nondisabled</i> (A_2)
<i>Independent Variable</i> Speech Type (B)		
Computer-synthesized speech (B_1)	$n = 48$	$n = 47$
Natural speech (B_2)	$n = 47$	$n = 47$

Diagram 5.2 2 × 2 Factorial Posttest Design (Between-Subjects)

Group	Treatment	Posttest
1	$X_{A_1B_1}$	O_1
2	$X_{A_1B_2}$	O_1
3	$X_{A_2B_1}$	O_1
4	$X_{A_2B_2}$	O_1

Time ►

Example for Diagram 5.2

Sidani, Y. M. (2007). Perceptions of leader transformational ability: The role of leader speech and follower self-esteem. *Journal of Management Development*, 26(8), 710–722.

Research Questions

- *Main effect:* Are individuals with low self-esteem more prone to the transformational message? Do leader speeches characterized by a good use of rhetorical devices lead to higher levels of attributed transformational abilities?
- *Interaction effect:* Does the interaction of the transformational leader style with inspirational speech lead to greater attributions of transformational leadership than when either or both of these components are absent? Is the same true for transactional leadership approach? Does the interaction of the nontransformational leader style with noninspirational speech lead to lower attributions of transformational leadership compared to other interactions?

Procedures: Participants were students enrolled in undergraduate business courses in a North American university. The participants were randomly assigned to one of four groups: (a) inspirational speech by a transformational leader ($n = 81$), (b) noninspirational speech by a transformational leader ($n = 77$), (c) inspirational speech by a transactional leader ($n = 76$), and (d) noninspirational speech by a transactional leader ($n = 79$).

Leadership style was manipulated by developing two leadership profiles describing the differences between transactional and transformational leaders. Next, participants were presented either with a noninspirational speech or an inspirational speech. All participants completed posttests on the Multifactor Leadership Questionnaire.

Design: Experimental research using a between-subjects approach 2×2 factorial posttest design

Recommended Parametric Analysis: 2-way between-subjects ANOVA or 2-way MANOVA (appropriate follow-up analysis plus descriptive statistics and effect-size calculations should be included)

<i>Assignment</i>	<i>Group</i>	<i>Factor Leader Type (A)</i>	<i>Factor Speech Type (B)</i>	<i>Posttest</i>
R	1	Transformational (A_1)	Inspirational speech (B_1)	Multifactor Leadership Questionnaire
R	2	Transformational (A_1)	Noninspirational speech (B_2)	Multifactor Leadership Questionnaire
R	3	Transactional (A_2)	Inspirational speech (B_1)	Multifactor Leadership Questionnaire
R	4	Transactional (A_2)	Noninspirational speech (B_2)	Multifactor Leadership Questionnaire
Time ►				

Note: The authors refer to the design as a randomized-factorial design, which means the order of the treatments was randomized (i.e., counterbalanced) for each participant to control for sequencing effects.

<i>Independent Variable</i> Speech Type (B)	<i>Independent Variable</i> Leader Type (A)	
	<i>Transformational (A_1)</i>	<i>Transactional (A_2)</i>
Inspirational (B_1)	$n = 81$	$n = 76$
Noninspirational (B_2)	$n = 77$	$n = 79$

Diagram 5.3

2 × 2 Factorial Pretest and Posttest Design
(Within-Subjects)

Group	Pretest	Treatment	Posttest
1	O ₁	X _{A1B1}	O ₂
1	O ₁	X _{A1B2}	O ₂
1	O ₁	X _{A2B1}	O ₂
1	O ₁	X _{A2B2}	O ₂

Time ►

Example for Diagram 5.3

Lopez, E. N., Drobles, D. J., Thompson, K. J., & Brandon, T. H. (2008). Effects of body image challenge on smoking motivation among college females. *Health Psychology, 27*(3), 243–251.

Research Questions: Does body dissatisfaction lead to a situational increase in smoking motivation among young college females? Do individuals with greater trait body dissatisfaction show the greatest impact of the body image challenge on smoking urges?

Procedures: Sixty-two participants were recruited to participate in the study. Participants were exposed to pairs of cues associated with each factor simultaneously on a split screen. Thus, the four conditions were thin models/smoking cues, thin models/neutral cues, neutral cues/smoking cues, and neutral cues/neutral cues. Initially, participants completed measures assessing basic demographics, Smoking Status Questionnaire (SSQ), and the Short-Smoking Consequences Questionnaire (Short-SCQ). The order of the treatment conditions were randomized for all the participants. The in-test and posttest measures included the Eating Disorders Examination-Questionnaire Version-Weight Concern subscale (EDEQ-WC), the Urge to Smoke Visual Analogue Scale (Smoke VAS), and the Weight Dissatisfaction Visual Analogue Scale (Weight Dissatisfaction VAS).

Design: Experimental research using a within-subjects approach 2 × 2 factorial pretest and posttest design

<i>Group</i>	<i>Pretest</i>	<i>Factor Body Image (A)</i>	<i>Factor Smoking Cue (B)</i>	<i>In-Test/Posttest</i>
1	SSQ, Short-SCQ	Thin (A ₁)	Smoking (B ₁)	EDEQ-WC, Smoke VAS, Weight Dissatisfaction VAS
1	SSQ, Short-SCQ	Thin (A ₁)	Neutral (B ₂)	EDEQ-WC, Smoke VAS, Weight Dissatisfaction VAS
1	SSQ, Short-SCQ	Neutral (A ₂)	Smoking (B ₁)	EDEQ-WC, Smoke VAS, Weight Dissatisfaction VAS
1	SSQ, Short-SCQ	Neutral (A ₂)	Neutral (B ₂)	EDEQ-WC, Smoke VAS, Weight Dissatisfaction VAS
Time ►				

2 × 2 (N = 62)	Independent Variable Body Image Manipulation (A)	
Independent Variable Smoking Cue (B)	<i>Thin (A₁)</i>	<i>Neutral (A₂)</i>
Smoking (B ₁)	n = 62	n = 62
Neutral (B ₂)	n = 62	n = 62

Note: Each cell represents the same participants (i.e., within-subjects).

Diagram 5.4 3 × 2 Factorial Posttest Design (Between-Subjects)

<i>Group</i>	<i>Treatment</i>	<i>Posttest</i>
1	X _{A1B1}	O ₁
2	X _{A2B1}	O ₁
3	X _{A3B1}	O ₁
4	X _{A1B2}	O ₁
5	X _{A2B2}	O ₁
6	X _{A3B2}	O ₁
Time ►		

Recommended Parametric Analysis: 2-way RM-ANOVA or 2-way MANOVA (appropriate follow-up analysis plus descriptive statistics and effect-size calculations should be included)

Example for Diagram 5.4

Gier, V. S., Kreiner, D. S., & Natz-Gonzalez, A. (2009). Harmful effects of preexisting inappropriate highlighting on reading comprehension and metacognitive accuracy. *The Journal of General Psychology*, 136(3), 287–300.

Research Questions

- *Main effect:* Does inappropriate highlighting reduce performance on a comprehension test in comparison with no highlighting and appropriate highlighting? Does inappropriate highlighting result in reduced ability to monitor comprehension and test performance? Does more difficult reading material result in lower comprehension and metacognitive accuracy?
- *Interaction effect:* Do metacognitive accuracy and comprehension scores differ with respect to text difficulty as a result of no highlighting, appropriate highlighting, and inappropriate highlighting?

Procedures: Participants were 180 undergraduate students from a southeastern university in the United States. The researchers manipulated text difficulty (two levels) and type of highlighting (three levels). Participants read texts that were either (a) not highlighted (control group), (b) appropriately highlighted (i.e., the sentences relevant to the comprehension questions were highlighted), or (c) inappropriately highlighted (i.e., the sentences highlighted were not relevant to the comprehension questions). Participants were randomly assigned to one of the six conditions. Immediately after reading each passage, participants provided metacomprehension ratings by recording how well they thought they had comprehended the passage. Then, participants responded to the comprehension test, which consisted of six multiple-choice questions. After responding to the comprehension questions, participants provided confidence ratings that their answers were correct. The researchers measured and scored all of the metacomprehension visual analogue and comprehension scales.

Design: Experimental research using a between-subjects approach 3×2 factorial posttest design

Recommended Parametric Analysis: 2-way between-subjects ANOVA or 2-way MANOVA (appropriate follow-up analysis plus descriptive statistics and effect-size calculations should be included)

<i>Assignment</i>	<i>Group</i>	<i>Factor Highlighted (A)</i>	<i>Factor Text Difficulty (B)</i>	<i>Posttest</i>
R	1	None (A_1)	Low (B_1)	Metacomprehension scale, comprehension test, confidence
R	2	Appropriate (A_2)	Low (B_1)	Metacomprehension scale, comprehension test, confidence
R	3	Inappropriate (A_3)	Low (B_1)	Metacomprehension scale, comprehension test, confidence
R	4	None (A_1)	High (B_2)	Metacomprehension scale, comprehension test, confidence
R	5	Appropriate (A_2)	High (B_2)	Metacomprehension scale, comprehension test, confidence
R	6	Inappropriate (A_3)	High (B_2)	Metacomprehension scale, comprehension test, confidence
Time ►				

2×3 ($N = 180$)	<i>Independent Variable</i> Highlighted Reading Passages (A)		
<i>Independent Variable</i> Text Difficulty (B)	<i>None (A_1)</i>	<i>Appropriate (A_2)</i>	<i>Inappropriate (A_3)</i>
Low (B_1)	$n = 30$	$n = 30$	$n = 30$
High (B_2)	$n = 30$	$n = 30$	$n = 30$

Diagram 5.5 $2 \times 2 \times 2$ Factorial Posttest Design (Between-Subjects)

Group	Treatment	Posttest		Group	Treatment	Posttest
1	$X_{A_1B_1C_1}$	O_1		5	$X_{A_1B_1C_2}$	O_1
2	$X_{A_1B_2C_1}$	O_1		6	$X_{A_1B_2C_2}$	O_1
3	$X_{A_2B_1C_1}$	O_1		7	$X_{A_2B_1C_2}$	O_1
4	$X_{A_2B_2C_1}$	O_1		8	$X_{A_2B_2C_2}$	O_1
Time ►				Time ►		

Note: For ease of reading, Groups 1–4 and 5–8 have been placed side by side, but (for example) Group 1 and Group 5 are not related.

Example for Diagram 5.5

Baylor, A. L., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior*, 25, 450–457.

Research Questions

- *Main effect*: What is the effect of types of instruction, deictic gestures, and types of facial expressions on student perceptions of pedagogical agent persona, attitude toward content, and learning?
- *Interaction effect*: Is there an interaction effect between the type of instruction and agent nonverbal behavior?

Procedures: A total of 236 undergraduate students participated in this study. The participants were randomly assigned to one of eight treatment conditions (roughly $n = 29$ per cell). The three major factors (independent variables) were each portioned into two levels: type of instruction (procedural, attitudinal), deictic gestures (presence, absence), and facial expressions (presence, absence). The two major modules developed were the procedural and attitudinal. The participants were exposed to the conditions during a computer literacy course. Following the exposure to the modules and conditions, the participants were assessed in three different areas: perceptions of agent persona, attitudes, and learning. The perceptions were assessed with the Agent Persona Instrument (API).

Design: Experimental research using a between-subjects approach $2 \times 2 \times 2$ factorial posttest design

Recommended Parametric Analysis: 3-way between-subjects ANOVA or 3-way MANOVA (appropriate follow-up analysis plus descriptive statistics and effect-size calculations should be included)

<i>Assignment</i>	<i>Group</i>	<i>Factor Instruction Type (A)</i>	<i>Factor Deictic Gesture (B)</i>	<i>Factor Facial Expression (C)</i>	<i>Posttest</i>
R	1	Procedural (A ₁)	Presence (B ₁)	Presence (C ₁)	API, attitudes, learning
R	2	Procedural (A ₁)	Absence (B ₂)	Presence (C ₁)	API, attitudes, learning
R	3	Attitudinal (A ₂)	Presence (B ₁)	Presence (C ₁)	API, attitudes, learning
R	4	Attitudinal (A ₂)	Absence (B ₂)	Presence (C ₁)	API, attitudes, learning
R	5	Procedural (A ₁)	Presence (B ₁)	Absence (C ₂)	API, attitudes, learning
R	6	Procedural (A ₁)	Absence (B ₂)	Absence (C ₂)	API, attitudes, learning
R	7	Attitudinal (A ₂)	Presence (B ₁)	Absence (C ₂)	API, attitudes, learning
R	8	Attitudinal (A ₂)	Absence (B ₂)	Absence (C ₂)	API, attitudes, learning
Time ►					

$2 \times 2 \times 2$	<i>Independent Variable (IV)</i> Instruction Type (A)		N = 236
<i>IV</i> <i>Deictic Gesture (B)</i>	<i>Procedural Module (A₁)</i>	<i>Attitudinal Module (A₂)</i>	<i>IV</i> <i>Facial Expression (C)</i>
Presence (B ₁)	n = 29	n = 30	Presence C ₁
Absence (B ₂)	n = 29	n = 30	Absence C ₁
Presence (B ₁)	n = 29	n = 30	Presence C ₂
Absence (B ₂)	n = 29	n = 30	Absence C ₂

Diagram 5.6 $2 \times 3 \times 2$ Factorial Posttest Design (Between-Subjects)

Group	Treatment	Posttest	Group	Treatment	Posttest
1	X_{A1B1C1}	O_1	7	X_{A1B1C2}	O_1
2	X_{A1B2C1}	O_1	8	X_{A1B2C2}	O_1
3	X_{A1B3C1}	O_1	9	X_{A1B3C2}	O_1
4	X_{A2B1C1}	O_1	10	X_{A2B1C2}	O_1
5	X_{A2B2C1}	O_1	11	X_{A2B2C2}	O_1
6	X_{A2B3C1}	O_1	12	X_{A2B3C2}	O_1
Time ►			Time ►		

Note: For ease of reading, Groups 1–6 and 7–12 were placed side by side, but (for example) Group 1 and Group 7 are not related.

Example for Diagram 5.6

Young, P., & Miller-Smith, K. (2006). Effects of state mandated policy (site-based councils) and of potential role incumbents on teacher screening decisions in high and low performing schools. *Education Policy and Analysis Archives*, 14(7), 1–21.

Research Questions

- *Main effect:* What are the effects of state-mandated policy and the potential role of incumbents on teacher screening in high- and low-performing schools?
- *Interaction effect:* Do legislative action, the role of the evaluator, and accountability differ with respect to perceived skill levels and the willingness to select a candidate?

Procedures: Public school administrators from two neighboring states participated in this study. A total of 624 participants were randomly selected from these locations. Three factors (independent variables) were evaluated in this study. The first was the state-legislative process, which was broken down into two levels: site-based and nonsite-based. The next factor was the role of the evaluator within the teacher-screening process. The screening

process (i.e., role of the evaluator) was portioned into principals, teachers, and parents. The third factor was the academic performance of public school districts and was broken down into high and low performers. All the participants were randomly assigned to 1 of 12 conditions. Each participant was assessed on two dependent variables. The first outcome variable measured the candidate's perceived skill level in relation to the job-related criteria. The second outcome variable measured the willingness that an evaluator (teacher, principal, or parent) would consider a candidate for an interview.

<i>Assignment</i>	<i>Group</i>	<i>Factor Legislative Action (A)</i>	<i>Factor Role of Evaluator (B)</i>	<i>Factor Accountability (C)</i>	<i>Posttest</i>
R	1	Site-based (A ₁)	Principal (B ₁)	Low (C ₁)	Perceived skill level, willingness to select candidate
R	2	Site-based (A ₁)	Teacher (B ₂)	Low (C ₁)	Perceived skill level, willingness to select candidate
R	3	Site-based (A ₁)	Parent (B ₃)	Low (C ₁)	Perceived skill level, willingness to select candidate
R	4	Nonsite-based (A ₂)	Principal (B ₁)	Low (C ₁)	Perceived skill level, willingness to select candidate
R	5	Nonsite-based (A ₂)	Teacher (B ₂)	Low (C ₁)	Perceived skill level, willingness to select candidate
R	6	Nonsite-based (A ₂)	Parent (B ₃)	Low (C ₁)	Perceived skill level, willingness to select candidate
R	7	Site-based (A ₁)	Principal (B ₁)	High (C ₂)	Perceived skill level, willingness to select candidate
R	8	Site-based (A ₁)	Teacher (B ₂)	High (C ₂)	Perceived skill level, willingness to select candidate
R	9	Site-based (A ₁)	Parent (B ₃)	High (C ₂)	Perceived skill level, willingness to select candidate
R	10	Nonsite-based (A ₂)	Principal (B ₁)	High (C ₂)	Perceived skill level, willingness to select candidate
R	11	Nonsite-based (A ₂)	Teacher (B ₂)	High (C ₂)	Perceived skill level, willingness to select candidate
R	12	Nonsite-based (A ₂)	Parent (B ₃)	High (C ₂)	Perceived skill level, willingness to select candidate
Time ►					

$2 \times 3 \times 2$	IV Role of Evaluator (B)			N = 624
IV Legislative Action (A)	Principal (B ₁)	Teacher (B ₂)	Parent (B ₃)	IV Accountability (C)
Site-based (A ₁)	n = 52	n = 52	n = 52	Low (C ₁)
Nonsite-based (A ₂)	n = 52	n = 52	n = 52	Low (C ₁)
Site-based (A ₁)	n = 52	n = 52	n = 52	High (C ₂)
Nonsite-based (A ₂)	n = 52	n = 52	n = 52	High (C ₂)

Diagram 5.7 $2 \times 2 \times 2$ Mixed-Factorial Design (Mixed-Subjects)

Group	Treatment	Posttest	Group	Treatment	Posttest
1	X _{A1B1 C1C2}	O ₁	5	X _{A1B1 C2C1}	O ₁
2	X _{A1B2 C1C2}	O ₁	6	X _{A1B2 C2C1}	O ₁
3	X _{A2B1 C1C2}	O ₁	7	X _{A2B1 C2C1}	O ₁
4	X _{A2B2 C1C2}	O ₁	8	X _{A2B2 C2C1}	O ₁

Time ►

Note: In this particular study, the third factor (Independent Variable: C1 and C2) was the variable treated as the within-subjects condition. Therefore, all eight groups were exposed to both levels of Factor C (1 and 2).

Design: Experimental research using a between-subjects approach $2 \times 3 \times 2$ factorial posttest design

Recommended Parametric Analysis: 3-way between-subjects ANOVA or 3-way MANOVA (appropriate follow-up analysis plus descriptive statistics and effect-size calculations should be included)

Example for Diagram 5.7

Guadagno, R. E., & Sagarin, B. J. (2010). Sex differences in jealousy: An evolutionary perspective on online infidelity. *Journal of Applied Social Psychology, 40*(10), 2636–2655.

Research Questions

- *Main effect:* What is the effect of infidelity context and type on perceptions of jealousy? Are there differences in regard to gender?
- *Interaction effect:* Is there an interaction effect between infidelity type and the context?

Procedures: A total of 332 undergraduate students ($n = 132$ males; $n = 200$ females) participated in this study. The participants were randomly assigned to one of eight treatment conditions (between 30 and 50 per cell). The three major factors (independent variables) were each portioned into two levels: infidelity context (online or conventional), gender (male or female), and infidelity type (sexual or emotional). Infidelity context and gender were between subjects, and infidelity type was within subjects. The participants were then exposed to one of two hypothetical scenarios, which included a series of questions related to emotional and sexual infidelity either in an online environment or in a conventional environment. Following the exposure, the participants were asked to complete a survey to assess their level of jealousy.

Design: Experimental research using a $2 \times 2 \times 2$ mixed-factorial design

Recommended Parametric Analysis: 3-way mixed ANOVA or 3-way MANOVA (appropriate follow-up analysis plus descriptive statistics and effect-size calculations should be included)

<i>Assignment</i>	<i>Group</i>	<i>Factor (BS) Infidelity Context (A)</i>	<i>Factor (BS) Gender (B)</i>	<i>Factor (WS) Infidelity Type (C)</i>	<i>Posttest</i>
R	1	Online (A_1)	Male (B_1)	Emotional (C_1) and sexual (C_2)	Test
R	2	Online (A_1)	Female (B_2)	Emotional (C_1) and sexual (C_2)	Test
R	3	Conventional (A_2)	Male (B_1)	Emotional (C_1) and sexual (C_2)	Test
R	4	Conventional (A_2)	Female (B_2)	Emotional (C_1) and sexual (C_2)	Test
R	5	Online (A_1)	Male (B_1)	Sexual (C_2) and emotional (C_1)	Test

Assignment	Group	Factor (BS) Infidelity Context (A)	Factor (BS) Gender (B)	Factor (WS) Infidelity Type (C)	Posttest
R	6	Online (A ₁)	Female (B ₂)	Sexual (C ₂) and emotional (C ₁)	Test
R	7	Conventional (A ₂)	Male (B ₁)	Sexual (C ₂) and emotional (C ₁)	Test
R	8	Conventional (A ₂)	Female (B ₂)	Sexual (C ₂) and emotional (C ₁)	Test

Time ►

Note: Factors A and B were treated as between subjects and Factor C was treated as within subjects (BS = Between subjects; WS = Within subjects).

2 × 2 × 2	Independent Variable (IV) Infidelity Context (A)		N = 332
IV Gender (B)	Online (A ₁)	Conventional (A ₂)	IV Infidelity Type (C)
Male (B ₁)	n = 33	n = 33	Emotional (C ₁) and sexual (C ₂)
Female (B ₂)	n = 50	n = 50	Emotional (C ₁) and sexual (C ₂)
Male (B ₁)	n = 33	n = 33	Emotional (C ₁) and sexual (C ₂)
Female (B ₂)	n = 50	n = 50	Emotional (C ₁) and sexual (C ₂)

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. Explain the major premise for using a factorial design.
2. What is the minimum number of IVs required for a factorial design? What is the minimum number of levels that should be included for each IV?
3. Considering threats to internal validity, is a factorial design considered a “strong” design?

Exercise

Develop a hypothetical research scenario that would necessitate the use of a **2 x 2 Factorial Design**. The research will be considered nonexperimental.

1. Identify the research scenario, including the relevant two independent variables and dependent variable(s). Identify the two levels of each independent variable.
2. Develop the appropriate primary research question to be associated with this design, including the research questions that will address main effects and interactions.
3. Discuss the sampling strategy and technique used to access the appropriate sample.
4. Identify the assignment technique to be used.
5. In accordance with the assignment technique, discuss the various control techniques that will be used with this specific design.
6. Discuss the major threats to validity associated with this design and type of research (experimental). How will these threats be addressed, based on the discussion of the control techniques in the previous question?
7. Briefly discuss any limitations associated with this research scenario and the specific design.

CHAPTER 6

SOLOMON N-GROUP DESIGN

The Solomon four-group design (Solomon, 1949) was developed specifically to combine the strengths of both types of between-subjects approaches (pretest only and the pretest and posttest design) as a means to minimize the weaknesses associated with using only one type. As a result, most of the major threats to internal validity (e.g., testing) and construct validity (e.g., pretest sensitization) are minimized. The inclusion of a control (or comparison) group to a research design can strengthen the internal validity and the overall validity of the findings. However, as noted earlier, there are strengths and costs in using between-subjects pretest and posttest control group designs compared to that of between-subjects posttest-only control group designs. The Solomon four-group design is an extension of the factorial design and is considered one of the strongest experimental designs, but its application in the social sciences is uncommon. Many investigators believe that logistical considerations (e.g., time, costs, number of participants, statistical analysis) are too much to overcome when applying this design. Although Solomon's original work did not include a sound statistical analysis for this design, researchers have attempted to offer statistical solutions and recommendations for power analysis (Sawilowsky, Kelley, Blair, & Markman, 1994; Walton Braver & Braver, 1988).

Originally, the Solomon four-group design was developed to include only four groups. Specifically, the four-group design includes one treatment (or factor; $k = 1$), with Group 1 receiving the treatment with a pretest and

posttest, Group 2 receiving the pretest and posttest with no treatment, Group 3 receiving the treatment and only a posttest, and finally Group 4 receiving only the posttest. This allows the researcher to assess the main effects, as well as interaction effects between the pretest and no-pretest conditions. However, it has been proposed that the original design can include more than one treatment, thus extending the design to six groups for 2-factor models or eight groups for 3-factor models (i.e., Solomon N -group design). These designs allow researchers to test the effects of more than one type of treatment intervention against one another. Therefore, the design can be referred to as a *Solomon four-, six-, or eight-group design*. We present examples of research that used a Solomon four-group design ($k = 1$), one example of a six-group design ($k = 2$), and one example of an eight-group design ($k = 3$).

Most common threats to internal validity are related, but not limited, to these designs:

Experimental. This design controls for all threats to internal validity except for Instrumentation.

Quasi-Experimental. Instrumentation and Selection Bias

We refer the reader to the following article for full explanations and recommended analyses regarding Solomon four-, six-, and eight-group designs:

Steyn, R. (2009). Re-designing the Solomon four-group: Can we improve on this exemplary model? *Design Principles and Practices: An International Journal*, 3(1), 1833–1874.

Diagram 6.1 Solomon Four-Group Design

Group	Pretest	Treatment	Posttest
1	O ₁	X _A	O ₂
2	O ₃	X _B	O ₄
3	—	X _A	O ₅
4	—	X _B	O ₆

Time ►

Note: It is highly recommended that random assignment be used when applying the Solomon N -group designs.

Example for Diagram 6.1

Probst, T. M. (2003). Exploring employee outcomes of organizational restructuring. *Group & Organization Management*, 28(3), 416–439.

Research Questions

- *Main effect:* Does job security, job satisfaction, commitment, physical and mental health decline, and turnover intention increase following the announcement and implementation of organizational restructuring? Do individuals who are affected by organizational restructuring report lower levels of job security, less job satisfaction, more negative affective reactions, greater intentions to quit, lower levels of physical and mental health, higher levels of role ambiguity, and higher levels of time pressure than individuals not affected by organizational restructuring?
- *Interaction effect:* The authors of this study did not explore interaction effects. A 2×2 factorial design would allow for the examination of the interactions within a Solomon four-group design. In this study, each independent variable has two levels (treatment and no-treatment; pretest and no-pretest). See the chart that follows for an example of a 2×2 factorial design for this study.

Procedures: A stratified random sample of 500 employees from five state agencies going through reorganization was selected. The stratification was based on whether the employee was affected by the reorganization. A total of 313 employees (63% of the sample) participated in the study. The sample was divided into two groups: those affected by the reorganization ($n = 147$) and those unaffected by the reorganization ($n = 166$). In addition, all participants were randomly assigned to either a pretest ($n = 126$) or no pretest ($n = 187$) group. Data were collected at two different time points: (a) immediately prior to the workplace reorganization announcement and (b) 6 months following the merger announcement. There were four different groups of participants: (a) pretested and affected by the reorganization, (b) pretested but unaffected by the reorganization, (c) affected but not pretested, and (d) unaffected and pretested. A survey assessing each of the variables (see Research Questions) of interest was administered prior to the merger announcement and 6 months into the reorganization.

Design: Experimental research using a between-subjects approach and a Solomon four-group design

Recommended Parametric Analysis: 2-way factorial ANOVA or maximum likelihood regression (appropriate descriptive statistics and effect-size calculations should be included)

<i>Assignment</i>	<i>Group</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
R	1 (<i>n</i> = 64)	Survey	Organizational restructuring	Survey
R	2 (<i>n</i> = 62)	Survey	—	Survey
R	3 (<i>n</i> = 83)	—	Organizational restructuring	Survey
R	4 (<i>n</i> = 104)	—	—	Survey

Time ►

<i>2 × 2 (n = 313)</i>	<i>Independent Variable Organizational Restructure (X)</i>	
	<i>Yes *</i>	<i>No</i>
<i>Independent Variable Pretest</i>		
Yes	<i>n</i> = 64 (<i>O</i> ₂)	<i>n</i> = 62 (<i>O</i> ₄)
No	<i>n</i> = 83 (<i>O</i> ₅)	<i>n</i> = 104 (<i>O</i> ₆)

Note: *O*₂, for example, represents the posttest observation for Group 1 from the design structure.

Diagram 6.2 Solomon Six-Group Design

<i>Group</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
1	<i>O</i> ₁	<i>X</i> _A	<i>O</i> ₂
2	<i>O</i> ₃	<i>X</i> _B	<i>O</i> ₄
3	—	<i>X</i> _A	<i>O</i> ₅
4	—	<i>X</i> _B	<i>O</i> ₆
5	<i>O</i> ₇	—	<i>O</i> ₈
6	—	—	<i>O</i> ₉

Time ►

Note: The study had two treatment conditions (fake-good [*X*_A] and fake-bad [*X*_B]) with one control and one pretest–posttest no treatment group serving both treatment conditions; thus, technically, it is a Solomon six-group design, although the authors refer to it as a four-group design.

Example for Diagram 6.2

Whitman, D. S., Van Rooy, D. L., Viswesvaran, C., & Alonso, A. (2008). The susceptibility of a mixed model measure of emotional intelligence to faking: A Solomon four-group design. *Psychology Science*, *50*(1), 44–63.

Research Questions

- *Main effect:* Do individual differences influence the extent to which emotional intelligence (EI) can be faked? To what extent do individual differences affect how much EI can be faked? Does a within-subjects design produce a higher effect size than a between-subjects design?
- *Interaction effect:* Was there an interaction effect between pretesting and faking instructions?

Procedures: The study included a sample of 300 undergraduate psychology students. Three measures were collected from different groups at different time points during the course of the study: (a) general mental ability (GMA), (b) personality (IPI), and (c) emotional intelligence (EIS). All participants were randomly assigned to one of six experimental conditions, and each group took the EIS, either (a) once under respond honest instructions and then once under fake-good instructions (Group 1); (b) once under respond honest instructions and then once under fake-bad instructions (Group 2); (c) only once (no pretest) under fake-good instructions (Group 3); (d) only once (no pretest) under fake-bad instructions (Group 4); (e) twice, both times under respond honest instructions (Group 5); or (f) one time (no pretest) under respond honest instructions (Group 6). Only one control group (Group 6) and only one no treatment pretest–posttest group (Group 5) were used for both the fake-good and fake-bad conditions.

Design: Experimental research using a within- and between-subjects approach and a Solomon six-group design

Recommended Parametric Analysis: 2-way factorial ANOVA, hierarchical multiple regression, or maximum likelihood regression (appropriate descriptive statistics and effect-size calculations should be included)

Assignment	Group	Pretest	Treatment	Posttest
R	1 (n = 50)	GMA, IPI, EIS	Fake-Good	GMA, IPI, EIS
R	2 (n = 50)	GMA, IPI, EIS	Fake-Bad	GMA, IPI, EIS
R	3 (n = 50)	—	Fake-Good	GMA, IPI, EIS
R	4 (n = 50)	—	Fake-Bad	GMA, IPI, EIS
R	5 (n = 50)	GMA, IPI, EIS	—	GMA, IPI, EIS
R	6 (n = 50)	—	—	GMA, IPI, EIS
Time ►				

Note: Groups 5 and 6 did not receive the treatment (i.e., were not told to fake-good or fake-bad) and completed the EIS with the instruction to respond honestly.

2 × 2 (n = 200)	Independent Variable Fake-Bad (X_B)	
	Yes	No
Independent Variable Pretest		
Yes	n = 50 (O_4)	n = 50 (O_8)
No	n = 50 (O_6)	n = 50 (O_9)

2 × 2 (n = 200)	Independent Variable Fake-Good (X_A)	
	Yes	No
Independent Variable Pretest		
Yes	n = 50 (O_2)	n = 50 (O_8)
No	n = 50 (O_5)	n = 50 (O_9)

Example for Diagram 6.3

McCarthy, A. M., & Tucker, M. L. (2002). Encouraging community service through service learning. *Journal of Management Education*, 26(6), 629–647.

Diagram 6.3 Solomon Eight-Group Design

Group	Pretest	Treatment	Posttest
1	O ₁	X _A	O ₂
2	O ₃	X _B	O ₄
3	O ₅	X _C	O ₆
4	—	X _A	O ₇
5	—	X _B	O ₈
6	—	X _C	O ₉
7	O ₁₀	—	O ₁₁
8	—	—	O ₁₂

Time ►

Research Questions

- *Main effect:* Does the participation in a service-learning project affect a student's intention to participate in community service? Does exposure to a community-service lecture affect a student's intention to participate in community service?
- *Interaction effect:* Was there an interaction effect between the service-learning project and the community-service lecture condition and the intention to participate in community service? Which of the three conditions (i.e., service learning, service lecture, or both combined) had the greatest effect on intention to participate in community service? Was there an interaction between the students' pre-intervention intention and the particular intervention they received and their postintervention intention to engage in community service?

Procedures: The sample included 437 college students enrolled in business classes at a state university. The research design employed eight groups: six treatment groups (two educational groups: each by themselves and in combination) and two control groups. The first treatment was a lecture by an instructor about the benefits of community service; the second treatment was a service-learning project that required students to perform tasks related to the content of the course for a local nonprofit organization. The

third treatment combined both the lecture and the service-learning project interventions. The dependent variable was the students' intention to participate in community service. This outcome variable was measured via a questionnaire that was administered as a pretest on the first day of class and as a posttest during the last week of class.

Design: Quasi-experimental research using a between-subjects approach and a Solomon eight-group design

Recommended Parametric Analysis: 2- or 3-way factorial ANOVA, hierarchical multiple regression, means analysis, or maximum likelihood regression (appropriate descriptive statistics and effect-size calculations should be included)

<i>Assignment</i>	<i>Group</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
NR	1	Questionnaire	Lecture	Questionnaire
NR	2	Questionnaire	Service Learning	Questionnaire
NR	3	Questionnaire	Both	Questionnaire
NR	4	—	Lecture	Questionnaire
NR	5	—	Service Learning	Questionnaire
NR	6	—	Both	Questionnaire
NR	7	Questionnaire	—	Questionnaire
NR	8	—	—	Questionnaire
Time ►				

2×2 ($n = 210$)	<i>Independent Variable Lecture (X_A)</i>	
	<i>Yes</i>	<i>No</i>
<i>Independent Variable Pretest</i>		
<i>Yes</i>	O_2	O_{11}
<i>No</i>	O_7	O_{12}

2×2 (n = 286)	Independent Variable Service Learning (X_B)	
Independent Variable Pretest	Yes	No
Yes	O ₄	O ₁₁
No	O ₈	O ₁₂

2×2 (n = 182)	Independent Variable Both (X_C)	
Independent Variable Pretest	Yes	No
Yes	O ₆	O ₁₁
No	O ₉	O ₁₂

**Reviewing the Content
and Testing Your Knowledge**

Discussion Points

1. What are the strengths to conducting a study using the Solomon N-group design? Can cause and effect be established from such a design?
2. What are the different types of designs embedded in the Solomon N-group design?
3. What are the common threats to internal validity of this design? Why?

Exercise

Develop a hypothetical research scenario that would necessitate the use of a **Solomon N-Group Design**. The research will be considered nonexperimental.

1. Identify the research scenario including the relevant independent variable and dependent variable.
2. Develop the appropriate primary research question to be associated with this design, including the research questions that will address main effects and interactions.
3. Discuss the sampling strategy and technique used to access the appropriate sample.
4. Identify the assignment technique to be used.
5. In accordance with the assignment technique, discuss the various control techniques that will be used with this specific design.
6. Discuss the major threats to validity associated with this design and type of research (experimental). How will these threats be addressed in accordance based on the discussion of the control techniques in the previous question?
7. Briefly discuss any limitations associated with this research scenario and the specific design.

CHAPTER 7

SINGLE-CASE APPROACH

The single-case approach is often referred to as the *single-participant* or *single-subject design*. In addition, some single-case approaches use more than one participant ($N = 1$) and are referred to as *small-n designs*, but the emphasis and unit of analysis remain on the single subject as reporting guidelines are regularly updated and produced (see Tate et al., in preparation). We remain consistent with our terminology and refer to these as single-case approaches and reserve the word *design* for the specific type of design defined within the approach. A single-case approach is used to demonstrate a form of experimental control with one participant (in some instances more than one participant). As seen in within-subject and between-subject approaches, the major contingencies required to qualify as a “true” experiment are randomization of conditions to participants (i.e., counterbalancing) or random assignment of participants to conditions. However, in single-case approaches, the participant serves as his or her own control, as well as serving in the treatment during which repeated measures are taken. More specifically, each condition is held constant and the independent variable is systematically withheld and reintroduced at various intervals as a means to study the outcome. The interval between the variable being withheld and reintroduced is based on theoretical and logistical considerations. A rule of thumb may be to consider equal intervals; however, there may be conditions that require washout periods, creating unequal intervals.

As a reminder, the treatment is also the same as a factor or intervention, and it is the independent variable. Although there are still debates concerning the number of experimental replications required to determine causation, as well as issues related to power, single-case approaches take a unique approach to experimentation. The threats to internal validity associated with the single-case approach are similar to those found in the within-subjects approach (e.g., sequencing effects), primarily because of the issues related to collecting repeated measures. In most cases, this approach meets the critical characteristics of experimental control (see Manolov, Solanas, Bulté, & Onghena, 2010, and Shadish, 2014a, for a review of robustness and power of randomization tests in A-B-A-B designs).

There are many forms, variations, and names of research designs for single-case approaches. We discuss four of the major designs here, with the understanding that this is not a comprehensive coverage of all the designs developed within this approach. The primary goal of the single-case approach is to measure the dependent variable and at the very minimum measure it against the presence and absence of the independent variable (treatment or intervention). Therefore, the design logic of a single-case approach starts with the baseline, which is designated as *A*, and then the treatment is designated as *B*. See Table 7.1 for the explanation of design notations that are unique to single-case approaches.

Table 7.1 Design Notations for Single-Case Approaches

<i>Design Notation</i>	<i>Design Element</i>
A	Baseline
B	Treatment _A
C	Treatment _B
O _n	Observations

Note: An observation (O_n) indicates multiple measures within each phase.

The most basic design within this approach is the A-B design (i.e., the dependent variable is measured during the baseline and then again during the treatment). Most single-case approach designs represent some variation and extension of the A-B design. It is important to note that, in order to qualify as an experiment, a researcher would, at a minimum, need to employ an A-B-A design (i.e., this is to establish that there is indeed a functional relationship between the independent and dependent variables).

There are many other variations of this design structure such as A-B-A-B, B-A-B, or A-B-C-A (C is used to represent a second treatment or independent variable). Any variation of the A-B design can be employed based solely on theoretical and logistical considerations.

When a researcher wants to study more than one treatment at a time, a multi-element design (also referred to as *multitreatment* or *alternating-treatment designs*) can be employed. This design requires rapid shifts between or within treatments to establish experimental control, and it allows an investigator to research two or more treatments (sometimes up to five or six). The third type of design within this approach is the multiple baseline design. While the A-B and multi-element require a withdrawal or reversal of conditions, the multiple baseline design requires no withdrawal or reversal (i.e., some treatments have carryover effects, so withdrawal or reversal is not theoretically appropriate). Specifically, two or more baselines are established, and the intervention is introduced at various points (usually across participants), but it is never removed. Most multiple baseline designs include more than one participant, but they may be used on a single participant applying the multiple baselines across multiple behaviors (as measured by the dependent variables). As previously noted, many of the single-case approach applications include more than one participant; however, each participant is analyzed individually.

Finally, there is a changing criterion design. Similar to the multiple baseline design, the changing criterion design allows for a gradual systematic manipulation of a targeted outcome and does not require a reversal or return to baseline phase as in the A-B design. This design is best applied when the researcher is interested in observing the stepwise increases of the targeted behavior.

We included three examples of the A-B design. Specifically, an A-B-A design, an A-B-A-B design, and an A-B-A-B-C-B-C design are presented. We also introduce one example of a changing criterion design and two multiple baseline designs (a 1-factor and a 2-factor design), which are forms of the basic A-B design.

The reader is referred to Dixon et al. (2009) to learn how to create graphs in Microsoft Excel for designs within the single-case approach. We also refer the reader to Shadish (2014a) for a review of the most recent issues regarding the analysis of the single-case approach such as modeling trend, modeling error covariances, computing standardized effect size estimates and assessing statistical power. In addition, we recommend the following article and book for a comprehensive overview of the single-case approach, specific forms of analysis for this approach, and software designed to analyze data from the family of A-B designs:

Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). London, England: Routledge.

Shadish, W. R. (2014b). Statistical analyses of single-case designs: The shape of things to come. *Current Directions in Psychological Science*, 23, 139–146.

Diagram 7.1 A-B-A Design

Case	Baseline A	Treatment B	Baseline A
1	O ₁ . . . O ₄	O ₁ . . . O ₄	O ₁ . . . O ₅
Time ►			

Example for Diagram 7.1

Haydon, T., Mancil, G. R., & Van Loan, C. (2009). Using opportunities to respond in a general education classroom: A case study. *Education and Treatment of Children*, 32(2), 267–278.

Research Question: What are the effects of a choral responding procedure and increased rate of delivering questions for an elementary student identified as at risk?

Procedures: An 11-year-old African American female student was the target participant for this study ($N = 1$). She was selected based on the fact that she displayed chronic disruptive behavior and an elevated score on the Systematic Screening for Behavior Disorders. The study took place in a science classroom. The classroom was not self-contained but considered departmentalized for the particular science class. The participant was measured on three separate dependent variables: (a) disruptive behavior, (b) correct response, and (c) on-task behavior. The treatment was identified as presenting a mode of questioning called *choral responding*. It consisted of an increased rate of questions and varied mode of questioning. The participant was observed during a baseline period when the teacher engaged in routine classroom instruction. Next, the choral responding technique was used; at this time point, the participant was again assessed on the three outcome measures. Last, the class returned to a routine lecture, and the participant was assessed again on the three outcome measures.

Design: Experimental research using a single-case approach and an A-B-A design

Recommended Statistical Analysis: Time-series analysis, autocorrelation, chi-square, or descriptive statistics (frequency, duration, latency, interresponse time, and celeration)

Case	Baseline (A) Four Observations	Treatment (B) Four Observations	Baseline (A) Five Observations
1	Disruptive behavior, correct response, on-task behavior	Disruptive behavior, correct response, on-task behavior	Disruptive behavior, correct response, on-task behavior
Time ►			

Diagram 7.2 A-B-A-B Design

Case	Baseline A	Treatment B	Baseline A	Treatment B
1	O ₁ . . . O ₁₀	O ₁ . . . O ₁₀	O ₁ . . . O ₁₃	O ₁ . . . O ₁₀
2	O ₁ . . . O ₁₀	O ₁ . . . O ₁₀	O ₁ . . . O ₁₃	O ₁ . . . O ₁₀
3	O ₁ . . . O ₁₀	O ₁ . . . O ₁₀	O ₁ . . . O ₁₃	O ₁ . . . O ₁₀
Time ►				

Note: Although this example includes a total of three subjects ($N = 3$), the emphasis remains only on the single subject, and the data between cases are not aggregated.

Example for Diagram 7.2

Bernard, R. S., Cohen, L. L., & Moffett, K. (2008). A token economy for exercise adherence in pediatric cystic fibrosis: A single-subject analysis. *Journal of Pediatric Psychology, 34*(4), 354–365.

Research Question: What are the effects of a token economy on exercise adherence in children with cystic fibrosis?

Procedures: The study included three participants ($N = 3$) between the ages of 8 and 12. Initially for the baseline phase of the study, the participants were asked to complete the Children’s OMNI Scale of Perceived Exertion and an exercise diary. The diary required them to track minutes of exercise

per day. The treatment phase consisted of two parts, which were the training and implementation. The participants were trained how to properly exercise and ensure activity levels stayed at an appropriate level. The token economy was built in as part of the treatment so that whenever the participants exercised they would receive a “token” (i.e., small prize or money). The participants and their parents were also properly trained on dietary needs. The parents were asked to continue with the token economy during the treatment phase for at least 1 week. After the treatment phase, the token economy was removed for approximately 10 days, then returned for the next week, then removed again for the final phase. The participants completed the OMNI scale and exercise diary throughout the process.

Design: Experimental research using a single-case approach and an A-B-A-B design

Recommended Statistical Analysis: Time-series analysis, autocorrelation, chi-square, or descriptive statistics (frequency, duration, latency, interresponse time, and celeration)

Case	Baseline (A) 10 Observations	Treatment (B) ~10 Observations	Baseline (A) 13 Observations	Treatment (B) ~10 Observations
1	OMNI Scale, exercise diary	OMNI Scale, exercise diary	OMNI Scale, exercise diary	OMNI Scale, exercise diary
2	OMNI Scale, exercise diary	OMNI Scale, exercise diary	OMNI Scale, exercise diary	OMNI Scale, exercise diary
3	OMNI Scale, exercise diary	OMNI Scale, exercise diary	OMNI Scale, exercise diary	OMNI Scale, exercise diary
Time ►				

Diagram 7.3 A-B-A-B-C-B-C Design

Case	Baseline A	Treatment _A B	Baseline A	Treatment _A B	Treatment _B C	Treatment _A B	Treatment _B C
1	O _n	O _n	O _n	O _n	O _n	O _n	O _n
2	O _n	O _n	O _n	O _n	O _n	O _n	O _n
3	O _n	O _n	O _n	O _n	O _n	O _n	O _n
Time ►							

Note: Each O_n represents multiple measures (observations), sometimes referred to as sessions.

Example for Diagram 7.3

Mancil, G. R., Haydon, T., & Whitby, P. (2009). Differential effects of paper and computer-assisted Social Stories™ on inappropriate behavior in children with autism. *Focus on Autism and Other Developmental Disabilities*, 24(4), 205–215.

Research Question: What are the effects of using a social story on aberrant behaviors of children diagnosed with autism?

Procedures: Three elementary students ($N = 3$) diagnosed with autism between the ages of 6 and 9 participated in this study. The behaviors that were targeted and measured were pushing, grabbing, touching, and shoving. The first intervention (Treatment_a) was a computer-assisted social skills training (CASST), which included a social story told using a PowerPoint format. The story was developed to address the specific needs of autistic children. The second intervention (Treatment_b) included the same social story, but it was presented using a paper format. For the initial baseline condition, teachers were asked to measure aberrant behaviors (e.g., pushing, grabbing) using an event-recording system. Next, the participants were introduced to the storytelling intervention using the CASST. Another baseline was measured, and then the storytelling format using CASST was again administered. Finally, the participants were exposed to the storytelling intervention with the use of paper. A return to the original intervention was presented (CASST), and then participants were again exposed to the paper storytelling intervention. Throughout the process, teachers recorded the target behavior with the event-recording system.

Design: Experimental research using a single-case approach and an A-B-A-B-C-B-C design

Recommended Statistical Analysis: Time-series analysis, autocorrelation, chi-square, or descriptive statistics (frequency, duration, latency, interresponse time, and celeration)

Case	Baseline (A) Five Observations	Treatment _a (B) Five Observations	Baseline (A) Five Observations	Treatment _a (B) Six Observations	Treatment _b (C) Five Observations	Treatment _a (B) Six Observations	Treatment _b (C) Five Observations
1	Behavior	Behavior	Behavior	Behavior	Behavior	Behavior	Behavior
2	Behavior	Behavior	Behavior	Behavior	Behavior	Behavior	Behavior
3	Behavior	Behavior	Behavior	Behavior	Behavior	Behavior	Behavior
Time ►							

Diagram 7.4 Changing Criterion Design (A-B)

Case	Baseline (A)	Treatment (B)
1	$O_1 \dots O_5$	$O_1 \dots O_{20}$
2	$O_1 \dots O_5$	$O_1 \dots O_{20}$
3	$O_1 \dots O_5$	$O_1 \dots O_{20}$

Time ►

Note: Any number of observations can be made during the treatment phase until the designated criterion is met.

Example for Diagram 7.4

Ganz, J. B., & Flores, M. M. (2009). The effectiveness of direct instruction for teaching language to children with autism spectrum disorders: Identifying material. *Journal of Autism and Developmental Disorders*, 39, 75–83.

Research Question: What are the effects of a direct-instruction language program on students with autistic spectrum disorders and their oral skills?

Procedures: Three participants ($M_{\text{age}} = 10.5$; $N = 3$) diagnosed with autism participated in this study. The intervention for this study was a direct-instruction program developed from the Language for Learning Materials. The specific strand of instruction chosen was the identification of common materials. These included items such as shirts, pants, robes, napkins, and leather shoes. The researchers created language probes designed after the tasks of the direct instruction intervention in order to assess the correct identification of items developed from the differential materials. For the initial phase, baseline data were collected until each individual performed consistently. Next, the treatment phase was implemented. The treatment phase lasted at each criterion level until all three participants achieved the predetermined criterion.

Design: Quasi-experimental research using a single-case approach and changing criterion design

Recommended Statistical Analysis: Time-series analysis, autocorrelation, chi-square, or descriptive statistics (frequency, duration, latency, inter-response time, and celeration)

Case	Baseline (A) 5 Observations	Treatment (B) 20 Observations
1	Language probe	Language probe
2	Language probe	Language probe
3	Language probe	Language probe
Time ►		

Diagram 7.5 Multiple Baseline Design (A-B)

Case	Baseline A_1	Treatment B
1	$O_1 \dots O_5$	$O_1 \dots O_{24}$
Case	Baseline A_2	Treatment B
2	$O_1 \dots O_9$	$O_1 \dots O_{15}$
Case	Baseline A_3	Treatment B
3	$O_1 \dots O_{13}$	$O_1 \dots O_{16}$
Time ►		

Note: The baselines for A_1 , A_2 , and A_3 should vary in length. The vertical dashed line signifies the varied length for the baseline phase between cases.

Example for Diagram 7.5

Schoenfeld, N. A., & Mathur, S. R. (2009). Effects of cognitive-behavioral intervention on the school performance of students with emotional or behavioral disorders and anxiety. *Behavioral Disorders*, 34(4), 184–195.

Research Question: What are the effects of cognitive-behavioral intervention on academic engagement and anxiety in children diagnosed with emotional and behavioral disorders?

Procedures: Three students ($N = 3$) were recruited to participate in this study. Each participant attended a private school for students with emotional and behavioral disorders. Students' levels of anxiety were assessed

by their teacher using the Child Symptom Inventory IV (CSI-IV). Participants' levels of academic engagement and school-appropriate behavior were also assessed as outcome measures. The intervention was the FRIENDS program, which is a cognitive-behavioral curriculum designed to help anxious children with emotional and behavioral disorders. After the initial baseline phase, each participant was exposed to the treatment twice a week in 30-minute intervals for a total of 29 sessions for each participant, including the baseline phase.

Design: Quasi-experimental research using a single-case approach and a multiple baseline design

Recommended Statistical Analysis: Time-series analysis, autocorrelation, chi-square, or descriptive statistics (frequency, duration, latency, interresponse time, and celeration)

Case	Baseline (A) ~5 Observations	Treatment (B) ~24 Observations
1	CVI-IV, engagement, behavior	CVI-IV, engagement, behavior
Case	Baseline (A) ~9 Observations	Treatment (B) ~15 Observations
2	CVI-IV, engagement, behavior	CVI-IV, engagement, behavior
Case	Baseline (A) ~13 Observations	Treatment (B) ~16 Observations
3	CVI-IV, engagement, behavior	CVI-IV, engagement, behavior
Time ►		

Diagram 7.6 Multiple Baseline Design (A-B-C)

Case	Baseline A_1	Treatment $_A$ B	Treatment $_B$ C
1	$O_1 \dots O_n$	$O_1 \dots O_n$	$O_1 \dots O_n$
Case	Baseline A_2	Treatment $_A$ B	Treatment $_B$ C
2	$O_1 \dots O_n$	$O_1 \dots O_n$	$O_1 \dots O_n$
Time ►			

Note: Four cases were used in this study.

Example for Diagram 7.6

Lyon, A. R., Gershenson, R. A., Farahmand, F. K., Thaxter, P. J., Behling, S., & Budd, K. S. (2009). Effectiveness of teacher–child interaction training (TCIT) in a preschool setting. *Behavior Modification*, 33(6), 855–884.

Research Question: What is the impact of teacher–child interaction training on preschool teachers' positive attention and discipline skills as a means to enhance children's psychosocial functioning and prevent mental health problems?

Procedures: Four classrooms and 12 teachers participated in this study. There were 3 teachers associated with each classroom, and each classroom represented a case ($N = 4$). The study included four experimental phases: (a) baseline, (b) treatment_A, (c) treatment_B, and (d) follow-up. The first 2 to 4 weeks served as the baseline period in which teachers were observed during routine instructional activities with the use of the Dyadic Parent-Child Interaction Coding System (DPICS). Next, the child-directed interaction (CDI) program was implemented and the observations with the DPICS continued. Then, after about 10 days, the teacher-directed

Case	Baseline (A) ~3 Observations	Treatment (B) ~11 Observations	Treatment (C) ~10 Observations
1	DPICS	DPICS	DPICS
Case	Baseline (A) ~6 Observations	Treatment (B) ~9 Observations	Treatment (C) ~10 Observations
2	DPICS	DPICS	DPICS
Case	Baseline (A) ~9 Observations	Treatment (B) ~13 Observations	Treatment (C) ~6 Observations
3	DPICS	DPICS	DPICS
Case	Baseline (A) ~11 Observations	Treatment (B) ~13 Observations	Treatment (C) ~7 Observations
4	DPICS	DPICS	DPICS
Time ►			

Note: Each case represented one classroom and three teachers.

interaction (TDI) program was implemented. The DPICS observation continued through this phase as well. At the commencement of the CDI and TDI phases, the teachers completed a teacher satisfaction survey. Classes 1, 2, and 3 were observed a total of 28 times, and then Class 4 was observed a total of 31 times over the baseline and treatment phases.

Design: Experimental research using a single-case approach and a multiple baseline design

Recommended Statistical Analysis: Time-series analysis, autocorrelation, chi-square, or descriptive statistics (frequency, duration, latency, inter-response time, and celeration)

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. Is the single-case approach as “powerful” as a traditional or classical experimental design?
2. Discuss some applications for which the single case approach is best suited.
3. What are some of the general threats to internal validity of this design? Why?

Exercise

Develop a hypothetical research scenario that would necessitate the use of an **A-B-A Design**. The research will be considered nonexperimental.

1. Identify the research scenario, including the relevant independent variable and dependent variable.
2. Develop the appropriate primary research question to be associated with this design.
3. Discuss the sampling strategy and technique used to access the appropriate case.
4. Discuss the various control techniques that will be used with this specific design.

5. Discuss the major threats to validity associated with this design and type of research (experimental).
6. How will these threats be addressed in accordance, based on the discussion of the control techniques in the previous question?
7. Identify the number of observations for each phase and a rationale for the number of observations.
8. Briefly discuss any limitations associated with this research scenario and the specific design.

PART II

Quantitative Methods for Nonexperimental Research

This part includes two popular approaches to the quantitative method (nonexperimental only), followed by some of the associated basic designs (accompanied by brief descriptions of published studies that use the design). Visit the companion website at study.sagepub.com/edmonds2e to access valuable instructor and student resources. These resources include PowerPoint slides, discussion questions, class activities, SAGE journal articles, web resources, and online data sets.

Nonexperimental research is conducted when the researcher does not have direct control of the independent variables simply because their

Method	Quantitative						
▼	▼						
Research	Nonexperimental						
▼	▼						
Approach	Observational			Survey			
▼	▼						
Design	Posttest (one-group)	Ex post facto	Explanatory	Predictive	Cross-sectional	Longitudinal	

Note: Quantitative methods for nonexperimental research are shown here followed by the approach and then the design.

manifestations have already occurred. That is, nonexperimental research is used when the variables of interest cannot be controlled through the means of manipulation, inclusion, exclusion, or group assignment. However, the one form of control that can be used for nonexperimental research is through different types of statistical procedures. This form of control can include the matching or pairing of alternate forms of independent variables (e.g., gender, socioeconomic status) prior to conducting an analysis or through the application of regression-discontinuity approaches to existing data sets, causal-modeling techniques, or propensity-scoring methods (i.e., equating groups on known covariates and assigning to conditions). Although the research is considered nonexperimental, it does not imply that less value or meaning may be derived from the findings. Typically, causal relationships can only be inferred from experimental research; but considering the limitations and difficulty applying experimental research, findings from nonexperimental research (observational/correlational) can be a compelling indicator of cause and effect (e.g., the relationship between smoking cigarettes and lung cancer).

Much of our knowledge today regarding cause-and-effect relationships (from astronomy to epidemiology) is derived from the nonexperimental research of observational data. As previously noted, researchers believe that causal links can be drawn from examinations that are nonexperimental in nature, particularly with the application of causal-modeling techniques and propensity-scoring methods. The reader is referred to Rubin (2007, 2008) for more information regarding the approximation of observational approaches to experimental tenets, as well as Imbens and Rubin (2015) for statistical procedures designed to infer causal inference from observational data.

Again, in nonexperimental research, the researcher does not attempt to control or manipulate the actual conditions (i.e., independent variables); rather, control is exerted over levels of variables (through statistical procedures). Hence, technically speaking, the concept of internal validity does not apply to nonexperimental research. Nonexperimental research is primarily used to explain or predict relationships or to describe and measure the degree of association (relationship) among variables. However, unlike internal validity, issues related to external, construct, statistical conclusion validity should still be accounted for when conducting nonexperimental research, although these concepts were originally defined to address the issues related to the determination of cause-effect relations through experimental or quasi-experimental research.

We present a one-group posttest-only design, which is considered nonexperimental and one of the “weakest” designs presented in this book. We

also present one example of an ex post facto design. However, the two most common forms of quantitative methods and nonexperimental research are observational and survey approaches (sometimes referred to in texts as *correlational designs* and *descriptive research*, respectively), which are presented in this part. As a simple reminder, the critical difference between experimental and nonexperimental research is the concept of control of the independent variable(s).

CHAPTER 8

EX POST FACTO AND POSTTEST-ONLY DESIGNS

Another form of nonexperimental research is called the *ex post facto* (after the fact) design, sometimes referred to as *causal-comparative research* (the term *causal-comparative* is considered a misnomer). This widely misunderstood and underused design is an attempt at creating quasi-experimental research out of nonexperimental research. Ex post facto designs are used when the researcher cannot control the treatment variable (i.e., the treatment and control groups are selected after the treatment has occurred), and there are no pretest measures, while only a posttest is collected. Unlike all the designs in nonexperimental research, the ex post facto design is unique in that issues related to internal validity still should be considered when evaluating the outcomes. The major threats include history, selection bias, maturation, and attrition. Clearly, the most obvious threat is selection bias because groups are self-selected and nonrandomly assigned to conditions for a multitude of reasons. Therefore, researchers can implement some type of “control” for the selection-bias issue by using a post hoc *matched-grouping* technique. This allows the researcher to establish control over the variables of interest—that is, because the independent (treatment) variable is not manipulated, various levels of alternate independent variables (e.g., age or gender) can be statistically manipulated (controlled) and used as a means to include individuals

in the desired conditions. These alternate independent variables are sometimes referred to as *quasi-independent* variables because they are not subjected to the various control techniques (e.g., manipulation, elimination). See Diagram 8.1 for an example of an ex post facto design. The reader is referred to Giuffre (1997) and Spector (1981) for more details regarding ex post facto designs.

Diagram 8.1 Two-Group Ex Post Facto Design

Group	Treatment	Time delay	Posttest	Assignment	Group
1	X		O_1	M_A	1
				2	

Time ►

Note: M_A represents the matched-grouping criteria (i.e., statistical procedures) used as a means to include the desired participants in each condition. The assignment to conditions is conducted after the treatment has occurred.

The one-group posttest-only design (often referred to as the *one-shot case study*) is considered nonexperimental and a “weak” design. Although the one-group posttest-only design is nonexperimental, threats to internal validity still should be a consideration—that is, the major threats to internal validity associated with this design are what determine the limitations in assessing the outcome. The obvious threats are selection bias and special treatment. Because there is only one designated observation with no comparison groups or multiple observations within subjects, it is nearly impossible to rule out plausible alternative explanations (i.e., the identified cause cannot be determined to be the only explanation for the effect).

Diagram 8.2 Posttest Design (One-Group)

Group	Treatment	Posttest
1	X	O_1

Time ►

Note: Statistical procedures is the only form of control to be used in nonexperimental research. However, this design is unique in that the independent variable can also be controlled via elimination and manipulation. This is the only exception in nonexperimental research.

Example for Diagram 8.2

Morgan, B. J. (2001). Evaluation of an educational intervention for military tobacco users. *Military Medicine*, 166(12), 1094–2001.

Research Question: What is the short-term effect of a tobacco-hazard education intervention on tobacco use and intent to quit?

Procedures: A tobacco-hazard education intervention was developed and presented to military tobacco users. The presentation lasted approximately 1 hour with a follow-up question-and-answer period. One month after the intervention, participants were asked to complete a survey regarding tobacco use and their intent to quit.

Design: Nonexperimental research using a one-group posttest-only design

Recommended Statistical Analysis: Descriptive statistics, one-sample *t* test

Assignment	Group	Treatment	Posttest
NR	1 (N = 151)	Tobacco hazard education	Tobacco use, intent to quit
Time ►			

Diagram 8.3 Ex Post Facto Design

Group	Treatment		Posttest	Assignment	Group
1	X	<i>Time delay</i>	O ₁	M _A	1 ----- 2
Time ►					

Note: The current example included one treatment for both groups, but some applications of this design can include two separate treatments or a treatment and control.

Example for Diagram 8.3

Chapin, M. H., & Holbert, D. (2009). Differences in affect, life satisfaction, and depression between successfully and unsuccessfully rehabilitated persons with spinal cord injuries. *Rehabilitation Counseling Bulletin*, 53(1), 6–15.

Research Question: What are the differences between persons with spinal cord injuries who were successfully rehabilitated from those who were not in regard to affect, life satisfaction, and depression?

Procedures: A total of 67 individuals with spinal cord injuries participated in the study. The participants went through a series of rehabilitative programs from the Division of Vocational Rehabilitation Services (DVRS). Participants were then assigned (matched grouping) to conditions based on successful rehabilitation ($n = 36$) or unsuccessful rehabilitation ($n = 31$). Following the rehabilitation program, participants completed the following questionnaires: Positive and Negative Affect Scale (PANAS), Satisfaction With Life Scale (SWLS), and the Center for Epidemiological Studies–Depression Scale (CES-D).

Design: Nonexperimental research using an ex post facto design

Recommended Parametric Analysis: One-way ANOVA, MANOVA, or independent-samples t test (appropriate descriptive statistics and effect-size calculations should be included)

Group	Treatment	Posttest	Assignment	Group
N = 67	DVRS program	PANAS, SWLS, CES-D	Unsuccessful	1 ($n = 31$)
		PANAS, SWLS, CES-D	Successful	2 ($n = 36$)
Time ►				

Note: The matched-grouping assignment to each condition was based on the success of the rehabilitation program. However, for example, the researcher could have matched the groups on gender and only included males in one analysis and females in the second, although this approach would require a larger sample size.

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. Why would a researcher choose to conduct nonexperimental research? What are the strengths and weaknesses associated with this type of research? Can cause and effect be established via this type of research? Why or why not?

2. Although the *ex post facto* design and one group posttest-only design are nonexperimental research, how are these designs unique compared to the other designs designated as nonexperimental?

Exercise

Develop a hypothetical research scenario that would necessitate the use of an **Ex Post Facto Design**. The research will be considered nonexperimental.

1. Identify the research scenario, including the relevant independent variable (which would have already occurred) and dependent variable.
2. Develop the appropriate primary research question to be associated with this design.
3. Discuss how the statistical procedures control technique will be used as a means to assign participants to each group. Discuss the rationale for this technique.
4. *Ex post facto* is unique in that it is nonexperimental research, but aspects of internal validity still apply. Discuss the major threats to validity associated with this design and type of research (nonexperimental).
5. Briefly discuss any limitations associated with this research scenario and the specific design.

CHAPTER 9

OBSERVATIONAL APPROACH

The observational approach is considered a correlational approach to research. The researcher does not intervene or use experimental control (i.e., manipulation, elimination, inclusion, group or condition assignment), hence this is considered nonexperimental research. Investigators use this approach when they are interested in measuring the degree of association (i.e., relationship) between variables or to predict some outcome (criterion) based on the predictor variable(s). The only type of control that can be used for nonexperimental research and the observational approach is statistical procedures. Because statistical techniques are the only form of control to be applied to the observational approach, some researchers and authors tend to refer to the actual analysis as the research design. This is misleading and not accurate because the statistical analysis is a tool to be applied to summarize the data which is organized via the research questions and research design.

Using correlational analyses or regression analyses, researchers can measure the strength (magnitude) and direction of the relationship between variables or predict the influence one variable has on another. Researchers can also apply analysis to observational data that compares the mean differences of multiple groups. Some researchers prefer to use the word *explanation* instead of *prediction* (i.e., if a phenomenon can be explained, then it can be predicted, although a prediction does not infer

explanation). As previously noted, results from nonexperimental observational data oftentimes can provide a strong case for making causal inferences (e.g., the systematic observation of many data points over time, indicating that texting while driving increases the likelihood of getting into an accident). However, scientists should be cautious when making causal inferences based on nonexperimental observational data. Sample size and correlational inference is an important aspect to consider for the observational approach (see Anderson, Doherty, & Friedrich, 2008). The two most common designs within the observational approach are explanatory and predictive designs.

♦ EXPLANATORY DESIGN

Through correlational or regression analysis, investigators attempt to *explain* the degree of association between two (or more) variables (sometimes referred to as *relational research*). Analyses can also be used to compare means of multiple groups depending on the research objectives (e.g., ANOVA, *t* test). When applying this design, data are collected at one point in time (theoretically) from a single group. Data can also be collected from extant data sets (i.e., retrospective analysis); once the research questions are applied to the data, and the design is identified, the appropriate analyses can be employed. More advanced explanatory designs can include collecting data on multiple variables as a means to confirm the direct and indirect effects between the variables (i.e., researchers attempt to infer causation through the application of causal modeling and confirmatory factor analysis and are sometimes referred to as *single-stage* and *multistage models*). Various forms of regression such as multiple, canonical, and cluster analysis, along with structural equation modeling, are used to summarize the data from these advanced explanatory designs (see Kline, 2010, for more information on multistage models).

♦ PREDICTIVE DESIGN

The predictive design goes beyond the explanatory design in that it allows the researcher to anticipate or *predict* the outcome based on the analysis of the relationship between two or more variables. Within this design, at

least one variable is indicated as the predictor variable, and one variable is designated as the criterion or outcome variable. Advanced predictive designs can include multiple predictor variables, while requiring more advanced forms of regression analyses to summarize the data. *Time* is a factor built into this design, so the researcher will typically collect data on the predictor variable(s) at one point and then at a later point collect data on the criterion variable(s). Furthermore, data can be collected from extant data sets and the predictive design can thus be applied to set up the appropriate research questions and subsequent analyses, as long as the concept of time appropriately exists between the predictor and criterion variables.

Statistical Tools

Some statistical tools and example research questions follow that would warrant the use of the particular tool. These analyses can be applied to observational data, depending on the research objective, when there is no independent variable (IV) to be manipulated. These are simplified and general guidelines. As a reminder, various sources refer to IVs differently between various forms of research (refer to Chapter 1 for a further discussion on this topic). We will reserve the use of the term *IV* for experimental or quasi-experimental research, and for nonexperimental research, we will refer to the IV as the *predictor* variable and the outcome or dependent variable as the *criterion* variable.

Comparison of Means for the Explanatory Design (one criterion variable, otherwise known as *univariate comparison of means*)

t Test (*independent samples*). What are the differences between School A and School B on standardized test scores? The test score is the criterion variable.

ANOVA. What are the differences between Schools A, B, and C on standardized test scores?

ANCOVA. What are the differences between Schools A, B, and C on standardized test scores controlling for socioeconomic status (SES)? SES is considered the covariate.

Trend Analysis. What are the quality-of-life-ratings of working college students for freshmen, sophomores, juniors, and seniors? These students had jobs as they entered college and maintained work throughout all 4 years of their college span.

2-Way ANOVA. What are the differences between gender and SES on standardized test scores? The two predictors (IVs) used are gender (male and female) and SES (high and low), both broken down into two levels each.

RM-ANOVA. What are the differences in college freshmen's quality-of-life indicators at the start of college and then again 1, 3, 5, and 7 months later?

Predicting a Single Variable for the Predictive Design

Hierarchical Linear Regression. What is the level (or prediction) of perceived satisfaction of voters based on years of voting experience, hours devoted to studying candidates, and ideas of collectivism as reported by political science professors? The researchers are interested in predicting one aspect of satisfaction (criterion variable), which is broken down by three subscales in the assessment: individual, collective, and familial. The predictor variables are voting experience, number of hours devoted to studying candidates, and collectivism.

Multiple Regression. How do levels of exercise and mental training predict levels of HbA1c for patients with type 2 diabetes? The interaction between exercise and mental training will also be examined.

Mediation. To what extent is the predictive effect of exercise on levels of HbA1c for type 2 diabetics mediated by mental training?

Predicting More Than One Variable

Binary Logistic Regression. To what extent do gender, age, and entrance exam scores predict graduate school success? Success is the binary criterion variable and is conceptualized as successful or not successful.

Multinomial Logistic Regression. To what extent do narcissism, self-efficacy, and gender predict type of leadership style (autocratic, democratic, equal)? The criterion variable in this example (leadership style) is broken down into three levels: autocratic, democratic, equal.

There are many other forms and types of analyses that can be applied to observational data sets. Keep in mind that statistical guides and texts often refer to the variables as either *predictor* variables or *independent*

variables, and for the criterion variable, it is often referred to as the *outcome* or *dependent* variable. These analyses can be used for the explanatory or predictive designs based on the research question(s). Examples of other analysis that can be used, depending on access and type of data, are multi-level linear modeling, contextual modeling, mixed models, discriminant function analysis, and principal and exploratory component analysis.

Diagram 9.1

Explanatory Design

Variable	Observation
1	O ₁
2	O ₁

We refer the reader to the following books for further details regarding observational approaches and analyses:

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge, UK: Cambridge University Press.

Keith, T. Z. (2014). *Multiple regression and beyond* (2nd ed.). Boston, MA: Pearson.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2012). *Applied multivariate research: Design and interpretation* (2nd ed.). Thousand Oaks, CA: Sage.

Example for Diagram 9.1

Walker, C. O., & Greene, B. A. (2009). The relations between student motivational beliefs and cognitive engagement in high school. *Journal of Educational Research*, 102(6), 463–471.

Research Question: What is the association between classroom motivation variables and students' sense of belonging?

Procedures: Questionnaires were distributed to participants ($N = 249$) during the middle of the term and completed in English classes. Perceived instrumentality and self-efficacy items were taken from the Approaches to Learning Survey. Student and classroom-achievement goal orientations were measured using the Patterns of Adaptive Learning Survey. Sense of belonging was assessed using the Psychological Sense of School Membership scale.

Design: Nonexperimental research using an observational approach with an explanatory design

Recommended Statistical Analysis: Correlational analysis

<i>Variable</i>	<i>Observation</i>
Motivation	Self-efficacy, instrumentality, goals
Sense of belonging	Sense-of-belonging survey

Note: Motivation = Self-efficacy, perceived instrumentality, mastery goals, and performance-based goals.

Diagram 9.2 Predictive Design₁

<i>Variable</i>	<i>Observation</i>	<i>Observation</i>
Predictor	O ₁	—
Criterion	—	O ₁
Time ►		

Note: A designated period of time must elapse before data on the criterion are collected.

Example for Diagram 9.2

Erdogan, Y., Aydin, E., & Kabaca, T. (2008). Exploring the psychological predictors of programming achievement. *Journal of Instructional Psychology*, 35(3), 264–270.

Research Question: What mental factors significantly predict programming achievement?

Procedures: Forty-eight students ($N = 48$) completed four different measurement tools that served as the predictor variables: (a) KAI creativity scale, (b) Problem Solving Inventory (PSI), (c) General Skills Test Battery (GSTB), and (d) Computer Attitude Scale (CAS). The criterion variable was measured using the Programming Achievement Test (PAT) after students completed the programming language course designed to introduce them to basic concepts of structured programming.

Design: Nonexperimental research using an observational approach with a predictive design

Recommended Parametric Analysis: Regression analysis or discriminant analysis

Variable	Observation	Observation
Mental factors	KAI, PSI, GSTB, CAS	—
Programming achievement	—	PAT
Time ►		

Note: Observations from Time Points 1 and 2 are from the same participants.

Diagram 9.3 Predictive Design₂

Variable	Observation	Observation	Observation	Observation	Observation	Observation
Predictor	O ₁	—	O ₂	—	O ₃	—
Criterion	—	O ₁	—	O ₂	—	O ₃
Time ►						

Example for Diagram 9.3

Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in early school years. *Journal of Educational Psychology, 101*(3), 662–670.

Research Question: To what extent do factors of teacher expectations in the early school years predict future academic performance?

Procedures: Children from 10 sites were followed from first to fifth grade ($N = 2,892$). Two measures of children's academic abilities were collected in the spring of the children's first- ($n = 966$), third- ($n = 971$), and fifth-grade ($n = 955$) years: (a) teacher reports of classroom performance in reading and math and (b) children's scores on standardized measures. In the spring of the first, third, and fifth grades, children were administered two subtests from a standardized psychoeducational assessment. A discrepancy score between teacher report of child academic performance and children's observed performance on standardized tests was calculated to determine the congruency of teacher expectancy and academic performance.

Design: Nonexperimental research using an observational approach with a predictive design

Recommended Parametric Analysis: Regression analysis

<i>Variable</i>	<i>Observation</i>	<i>Observation</i>	<i>Observation</i>	<i>Observation</i>	<i>Observation</i>	<i>Observation</i>
Teacher expectations	Teacher expectancy	—	Teacher expectancy	—	Teacher expectancy	—
Child academic performance	—	Academic performance	—	Academic performance	—	Academic performance
Time ►						

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. What type of validity should be considered when using the observational approach?
2. Explain why the only type of control that can be applied to observational approach is statistical procedures.

Exercise

Develop a hypothetical research scenario that would necessitate the use of a **Predictive Design**. The research will be considered nonexperimental.

1. Identify the research scenario, including the relevant predictor variable and criterion variable.
2. Develop the appropriate primary research question to be associated with this design.
3. Describe why the observational approach and predictive design is the most appropriate methodology to be used considering the research scenario.
4. How is the concept of time being factored into this design?
5. What type of sampling procedure and sampling technique will be used to access the appropriate sample?
6. Discuss how the only form of control (statistical procedures) will be used in this scenario.
7. Pick several threats each from external, construct, and statistical conclusion validity, and discuss how they will be accounted for.
8. Briefly discuss any limitations associated with this research scenario and the specific design.

CHAPTER 10

SURVEY APPROACH

The most common form of nonexperimental research is the survey approach (sometimes referred to as *descriptive research*). Typically, investigators administer a survey to a randomly selected sample of individuals or, if possible, to an entire population (see Fowler, 2013). Random selection is a critical element to survey research in that generalization (external validity) is the primary goal of the findings (i.e., external validity is the focus; internal validity does not apply). However, it must be noted that construct validity and statistical conclusion validity do apply to the survey approach. Again, the concept of internal validity is concerned with the establishment of cause-effect relations, whereas the survey approach is not applied to determine cause and effect. Surveys are used to observe trends, attitudes, or opinions of the population of interest. Participants are usually selected from the population to discover the relative incidence, distribution, and interrelations of educational, sociological, behavioral, or psychological variables. Thus, it can be classified as quantitative and is often considered a variant of the observational approach.

Often, as noted, when applying the survey approach, the goal is to eventually generalize the findings to the entire population. To achieve this, some form of a probability sampling strategy should be employed when applying the survey approach. Many major news outlets conduct surveys on a regular basis, but they are considered nonscientific, and the findings are not expected to generalize to the greater population. Why? First, the scientific method was not applied to the scenario, and second, a nonprobability sampling strategy was used to collect the data (only viewers of that particular news source and users of the specific website participate).

♦ PRIMARY THREATS TO VALIDITY

External validity. Keep in mind that external validity, by definition, is the construct that is related to generalization. The major threat to external validity for the survey approach is *sample characteristics*. Sample characteristics are the extent to which the sample surveyed represents the identified population. To ensure this form of validity is not violated (assuming generalization is desired), then the appropriate probability sampling strategy should be employed. The following are the various types of probability sampling: simple random, cluster, stratified, systematic, and multistage.

Construct validity. The focus for the survey approach is seated in measurement. Therefore, it is vital that one can generalize the findings or results of a survey to the theory and primary inquiry posed by the researcher. Issues related to the survey instrument directly affect the validity of the outcome. The primary threats to construct validity for the survey approach are *reactivity to assessment* or *acquiescence response bias* and *timing of measurement*. Typically when participants know they are being surveyed, they may change or alter the way they respond to items on a survey, which is different than the way they truly feel. This is known as *reactivity to assessment* or more commonly known as *social desirability*. The second major threat is the timing of the measurement. The time in which the survey is administered can greatly affect the results due to numerous conditions outside the control of the researcher.

Statistical conclusion validity. Surveys and assessment tools should measure and gather the information that the instrument is purported to measure. This is fundamental to any form of research. And the greatest threat to statistical conclusion validity for the survey approach is *unreliability of the measures*. Clearly, if the measure is not reliable or suffers from inadequate levels of validity, then the result or outcome will be compromised.

Response rate. Response rates for the survey approach are always notoriously low. Often, researchers can expect a 15% to 20% return rate for external surveys, and internal surveys (workplace surveys) may exceed those percentages. Researchers are often limited in time and resources (e.g., money for incentives), which adds to the problem of securing adequate response rates. The quality of the response rate can directly affect the validity of the outcome. Fincham (2008) recommended a 60% response rate should be the general goal for most types of research, but researchers should strive for at least 80% for the survey approach when the intent is to generalize to the entire population. Again, these numbers are often unrealistic,

particularly for students conducting research for theses and dissertations. Mundia (2011) exemplified the threats to validity when social desirability and response rates are not secured, which should be noted in the potential limitations for all studies classified under the survey approach. As a result, in order to address “low” response rates, researchers should conduct non-response bias analysis. Drechsler (2015) and Thompson and Oliver (2012) offered recommendations for strategies to address nonresponse bias in the form of multiple and fixed-affect imputations.

The reader is referred to Lavrakas (2009) for an in-depth and comprehensive coverage of the survey approach and the open-access journal *Survey Methodology* (<http://www5.statcan.gc.ca/olc-cel/olc.action?objId=12-001-X&objType=2&lang=en&limit=0>).

CROSS-SECTIONAL DESIGN ♦

The cross-sectional design allows the researcher to collect data at one point in time. This design is one of the most common designs that media outlets use to present information of public opinion on political or social circumstances. The most common application of this design is to gather opinions or attitudes from one specific group. However, in many cases, the same instrument can be administered to different populations as a means to compare a group’s attitudes or opinions on the same variable. Basic descriptive statistical analyses are typically used to summarize data.

LONGITUDINAL DESIGN ♦

An extension of the cross-sectional design is the longitudinal design. This design allows the researcher to collect survey data over a designated period of time with the same or different samples within a population. Researchers can collect data using *trend* (identify a population to examine changes over time), *cohort* (identify a subpopulation based on specific characteristics), or *panel* (survey the same people over time) studies. Based on theoretical and logistical considerations, longitudinal designs can include any combination of data collected from cohorts, trends, and panels. Variations of this design include the *cohort-sequential design* and the *accelerated longitudinal design*. These designs allow the researcher to collect data from temporally related cohorts over time to determine the extent of the relation between the cohorts (see Prinzie & Onghena, 2005).

Diagram 10.1

Cross-Sectional Design

We refer the reader to the following book for further details regarding survey approaches:

Variable	Observation
1	O ₁

Fink, A. G. (2012). *How to conduct surveys: A step-by-step guide* (5th ed.). Thousand Oaks, CA: Sage.

Example for Diagram 10.1

Jones, M. A., Stratten, G., Reilly, T., & Unnithan, V. B. (2004). A school-based survey of recurrent non-specific low-back pain prevalence and consequences in children. *Health Education Research*, 19(3), 284–289.

Research Question: What evidence exists to demonstrate the prevalence and consequences of recurrent low-back pain in children?

Procedures: Questionnaires were issued to seven different schools (N = 500). A cross-sectional sample of 500 participants, boys (n = 249) and girls (n = 251), was collected. Participants were required to complete a questionnaire to assess their low-back pain history. The questionnaire was designed to identify lifetime prevalence, point prevalence, recurrent prevalence, and duration of the low-back pain.

Variable	Observations
Low-back pain	Low-back pain survey

Diagram 10.2

Longitudinal Design

Variable	Observation	Observation	Observation	Observation
1	O ₁	O ₂	O ₃	O ₄
2	O ₁	O ₂	O ₃	O ₄
3	O ₁	O ₂	O ₃	O ₄
4	O ₁	O ₂	O ₃	O ₄
5	O ₁	O ₂	O ₃	O ₄

Time ►

Note: Any number of variables and observations can be associated with this design.

Design: Nonexperimental research using a survey approach with a cross-sectional design

Recommended Statistical Analysis: Descriptive statistics

Example for Diagram 10.2

Steinfeld, C., Ellison, N. B., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology, 29*, 434–445.

Research Questions: How does the use of a social network among a college population change over time? What is the directionality of the relationship between social network use and development of bridging social capital? How does an individual's psychological well-being influence the relationship between social capital and social network site use?

Procedures: Survey data were collected from university students at two time points a year for 2 consecutive years. Initially, undergraduate students were

Variable	Observation	Observation	Observation	Observation
	Full Sample (n = 288)	Panel (n = 92)	Random Sample (n = 481)	Panel (n = 92)
Internet use	Internet use survey	Internet use survey	Internet use survey	Internet use survey
Social network use	Social network survey	Social network survey	Social network survey	Social network survey
Well-being	Well-being survey	Well-being survey	Well-being survey	Well-being survey
Self-esteem	Self-esteem survey	Self-esteem survey	Self-esteem survey	Self-esteem survey
Satisfaction	Satisfaction survey	Satisfaction survey	Satisfaction survey	Satisfaction survey
Time ►				

Note: Panel refers to the same participants that were surveyed from the full and random samples.

sent an e-mail invitation with a short description of the study, information about confidentiality, an incentive for participation, and a link to the survey. Participants were surveyed on general Internet use, social network use, psychological well-being, self-esteem, and satisfaction. As a follow-up to the first-year survey, in-depth interviews were conducted with 18 students primarily drawn from the initial sample.

Design: Nonexperimental research using a survey approach with a longitudinal design

Recommended Parametric Analysis: Descriptive statistics or correlational analysis

Reviewing the Content and Testing Your Knowledge

Discussion Points

1. What is the major difference between a cross-sectional and a longitudinal design?
2. What type of scenario would warrant the application of a longitudinal design overall a cross-sectional design?
3. Describe some of the threats to external validity that are common with the survey approach.

Exercise

Develop a hypothetical research scenario that would necessitate the use of a **Longitudinal Design**. The research will be considered nonexperimental.

1. Identify the research scenario, including the relevant variable(s).
2. Develop the appropriate primary research question to be associated with this design.
3. Describe why the longitudinal design is the most appropriate methodology to be used, considering the research scenario.
4. How is the concept of time being factored into this design?

5. What type of sampling procedure and sampling technique will be used to access the appropriate sample?
6. Discuss how the only form of control (statistical procedures) will be used in this scenario.
7. Pick several threats each from external, construct, and statistical conclusion validity, and discuss how they will be accounted for.
8. Briefly discuss any limitations associated with this research scenario and the specific design.

PART III

Qualitative Methods

This part includes four popular approaches to the qualitative method, followed by some of the associated basic designs (accompanied by brief descriptions of published studies that use the design). Visit the companion website at study.sagepub.com/edmonds2e to access valuable instructor and student resources. These resources include PowerPoint slides, discussion questions, class activities, SAGE journal articles, web resources, and online data sets.

The qualitative method represents a form of data collection and analysis, with a focus on understanding and an emphasis on meaning. Research under the qualitative method is considered emerging and nonexperimental. This method is often used to explore the “how” and “why” of systems and

<i>Method</i>	<i>Qualitative</i>					
▼	▼					
Research	Nonexperimental					
▼	▼					
Perspective	Grounded Theory	Ethnographic	Narrative		Phenomenology	
▼	▼					
Design	Systematic	Emerging	Constructivist	Realist	Critical	Case Study
	Descriptive	Explanatory	Existential	Transcendental	Hermeneutic	

Note: Qualitative methods for nonexperimental research are shown here, followed by the approach and then the design.

human behavior and what governs these behaviors. Specifically, it is a method for examining phenomena, predominantly using “words” for data. The qualitative process is generally inductive, although it can be abductive (an inductive-deductive cycle) and emerging. Qualitative researchers usually take a naturalistic approach to the world (i.e., studying things in their natural setting), while attempting to understand phenomena through the “voice” of the participants. Biases are accepted as part of the process (e.g., purposive sampling and the “researcher as instrument”). These biases make it critical that the researcher is fully aware of his or her ontological and epistemological stance that provides the framework for the research.

Behavior is generally studied as it occurs naturally, with no manipulation or control. The overarching aim of the qualitative method is to understand or interpret phenomena within the context of the meaning that people express, without attempting to infer causation or generalize (i.e., external validity) the results to other individuals or populations. However, as previously mentioned, the concept of external validity can have a place in qualitative methods such as *theory-focused* and *case-to-case generalization* (see Chenail, 2010, for a review on nonprobabilistic approaches to generalizability for qualitative methods), as well as improving issues related to validity (e.g., trustworthiness) and reliability through triangulation techniques (Golafshani, 2003).

Generally, the aim of the qualitative method is to reveal and understand phenomena within a particular context, without attempting to infer any type of causation. This is profoundly different from the nature of experimental research, which is designed to infer cause and effect; however, there is a unique technique used to infer causation from qualitative case studies known as *qualitative comparative analysis* (QCA). QCA, developed by Ragin (1997), allows researchers to combine the in-depth qualitative methodological strategies with a quantitative-oriented approach into a single framework. QCA is used when a researcher has data from a moderate number of cases (typically too many for a traditional cross-case qualitative analysis), then transcribes the data via qualitative analytic software and converts the data to dichotomous or ordinal data in preparation for the QCA analysis. At that point, the researcher attempts to draw causal inferences from the outcomes of the QCA analysis (see Rihoux & Lobe, 2009, for more on QCA applications).

Although we include many specific (named) designs used by qualitative researchers, it is important to discuss what we refer to as the *generic* qualitative design. The generic design for qualitative methods can borrow from any of the approaches and designs covered in this section. The generic design is a “catchall” approach, with qualitative data collection and

the successive qualitative analyses being the common thread between the various manifestations. The type of analysis for this approach is usually inductive, involving the organization of data into categories and the subsequent identification of patterns and relationships among these categories. Although the approaches to data collection and analytic style differ among these generic designs, the general process of qualitative data analysis is fairly standard (see Wertz et al., 2011).

THE CASE STUDY ♦

Many disciplines use various forms of the case study to examine an individual or phenomenon within a specified context. The approach and application of case study designs also can vary widely between various disciplines such as medicine, law, and the social sciences. However, in the social and behavioral sciences, case studies are often referred to as *uncontrolled* studies. Yin (2013) defined the case study as an empirical inquiry that investigates a phenomenon within its real-world context, when the boundaries between phenomena and context are not clearly evident, in which multiple data sources are used. Yin referred to the case study as a “method” as opposed to confining it to only an approach or a “tradition” within the various forms of qualitative research (e.g., Creswell, 2012). Generally, the focus of the case study is on developing a narrative or revealing a phenomenon based on an in-depth, real-time, or retrospective analysis of a case. Therefore, issues related to experimental control and internal validity are nonfactors within this approach. Although case studies do not infer causation and the results should not be generalized, the findings can provide rich insight toward phenomena and serve as support for theories and the generation of hypotheses. However, if desired, Yin does offer approaches and models for researchers interested in attempting to infer causation from case study designs (which differs from QCA analysis).

The emphasis in a case study is primarily the qualitative method; however, cross sections of quantitative data are usually collected as supplementary data throughout the analyses (see mixed method embedded case study design). The label of *case study* is often applied to many social science examinations as a catchall term, many times misapplying the concept (Malcolm, 2010). However, the case study design can be applied to any of the approaches within the qualitative method, such as the most commonly applied narrative and phenomenological approach in psychology (Singer & Bonalume, 2010a) or the ethnographic approach in education (Creswell, 2014).

Creswell took a different angle than Yin (2013) regarding the type and description of designs for the case study. Gall, Gall, and Borg (2007) succinctly described a case study “as (a) the in-depth study of (b) one or more instances of a phenomenon (c) in its real-life context that (d) reflects the perspective of the participants involved in the phenomenon” (p. 447).

Confusion does arise when authors use different terminology for similar constructs. These semantic differences can be seen in the work of Yin, who uniquely defined and applied the terms *holistic* and *embedded* (see Appendix B) differently than their traditional uses; for example, the term *embedded* has an entirely different meaning when used by Creswell. Another example of this is the term *case study design*, used within the qualitative method and most often associated with the ethnographic and phenomenological approaches. However, the case study can also be applied to the narrative approach and arguably any other approach within the qualitative method, as long as the “case” being explored is bound by time, place, person, or environment. When deciding to use a case study, we refer the reader to Yin’s (2004, 2012, 2013) books for a review of his unique and widely accepted approach to the case study. See Appendix B for a list of case study designs defined by Yin (2013) and Creswell (2014).

Recommended programs for qualitative data analyses: ATLAS/TI, The Ethnograph, HyperRESEARCH, NVivo, NUD*IST, SPSS Text Analysis for Surveys™

We refer the reader to the following books and book chapter for further details regarding qualitative methods:

Creswell, J. W. (2012). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Lincoln, Y. S., Linham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging influences., revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (4th ed.), pp. 191–215. Thousand Oaks, CA: Sage.

Wertz, F. J., Charmaz, K., McMullen, L. M., Josselson, R., Anderson, R., & Emalinda, M. (2011). *Five ways of doing qualitative analysis: Phenomenological psychology, grounded theory, discourse analysis, narrative research, and intuitive inquiry*. New York, NY: Guilford Press.

CHAPTER 11

GROUNDED THEORY PERSPECTIVE

The grounded theory approach was first developed by Glaser and Strauss (1967) as a way to generate a theory based on data that are systematically gathered and analyzed. In general, this is an inductive process in which the theoretical propositions are not presented a priori; rather, the theory emerges from the data that are being collected. However, this process often becomes abductive, with testing of the theory occurring as it emerges from the data. The emerging theory is constantly being compared to the evidence brought forth from new data that are analyzed, as in the “constant comparative method.” The use of *memoing* (i.e., the process of recording the personal thoughts and ideas of the researcher throughout the data collection procedures) is critical when using a systematic, emerging, or constructivist design. Qualitative researchers often use memoing to help make conceptual links between raw data and abstractions to better explain the phenomena being studied within its appropriate context. See Birks, Chapman, and Francis (2008) for an in-depth discussion of memo-writing techniques.

According to Corbin and Strauss (2015), a good grounded theory should (a) fit the phenomenon; (b) provide understanding; (c) provide generality, in that the theory includes extensive variation and is abstract enough to be applicable to a wide variety of contexts; and (d) provide control, in the sense of stating the conditions under which the theory applies and describing a reasonable basis for action.

♦ SYSTEMATIC DESIGN

- The systematic design is the most structured of the grounded theory approaches, with rigid procedures and a preconceived framework for categories. This design emphasizes theory verification based on the theory that is generated (i.e., inductive-deductive process). The design uses the three-stage coding method (open, axial, and selective) to help generate a visual depiction of a theory.

♦ EMERGING DESIGN

The emerging design is also a theory generation design; however, it is less prescriptive than the systematic design. This design allows the theory to emerge “naturally” from the data. The key components of this design are fit, work, relevance, and modifiability.

♦ CONSTRUCTIVIST DESIGN

The constructivist design further distances itself from the procedurally laden systematic design, stressing the role of the researcher as an active participant who interacts with the field being explored. Constructivist researchers are interested in the co-construction of knowledge between researcher and participant and embrace and explore the inherent biases within this interaction. This design recognizes that knowledge emerging from the data is not only “discovered” but also created. It is important to be cognizant of the assumptions brought to the investigation by the researcher. Also, one should be aware of the socially constructed meanings that occur during the collection of data and those socially constructed meanings that were in place prior to engaging with the participant.

WHEN TO USE GROUNDED THEORY

- To build/discover theory inductively
- To build/discover substantive and/or formal theory
- When there is little or no prior information on an area or phenomenon
- To study the microcosm of interaction

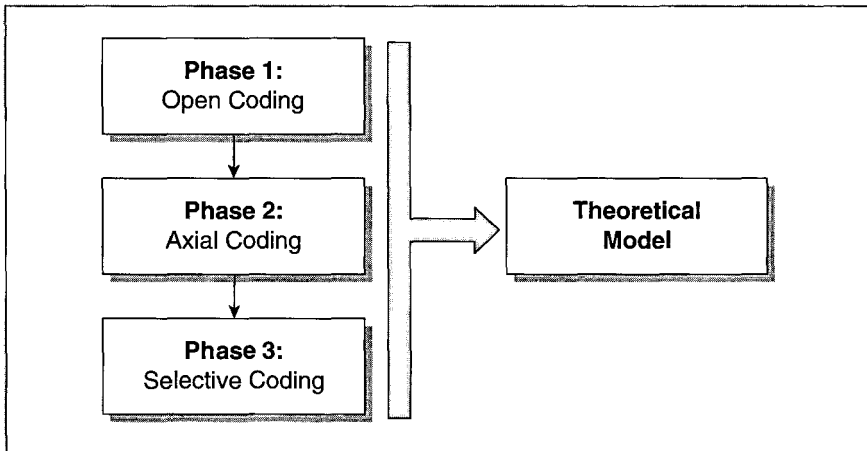
We refer the reader to the following books for further details regarding the grounded theory approach:

Charmaz, K. (2014). *Constructing grounded theory: A practical guide through qualitative analysis* (2nd ed.). London, England: Sage.

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Thousand Oaks, CA: Sage.

Glaser, B. G., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Hawthorne, NY: Aldine Transaction.

Figure 11.1 Systematic Design



Example for Figure 11.1

Han, G. S., & Davies, C. (2006). Ethnicity, health and medical care: Towards a critical realist analysis of general practice in the Korean community in Sydney. *Ethnicity and Health*, 11(4), 409–430.

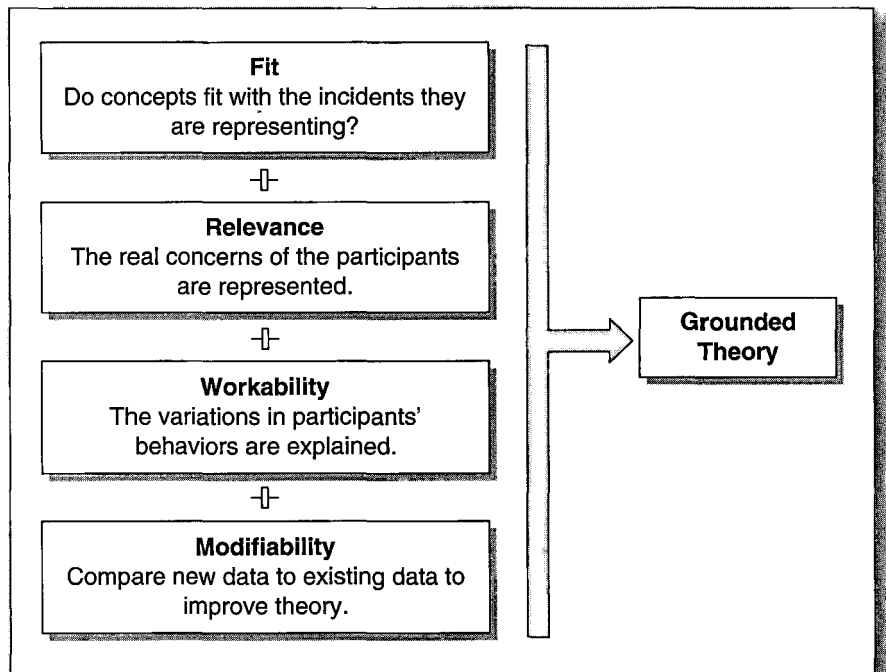
Research Question: What are the general practitioners' views on the health of Koreans and the complex process of providing and seeking effective and satisfactory medical care?

Procedures: This study investigated the use and provision of biomedicine among men on the basis of interview data from eight doctors. Semistructured interview schedules were prepared around the doctors'

views of (a) health, immigrant life, and health care use among Koreans from different socioeconomic backgrounds; (b) common ailments; (c) particular difficulties servicing fellow Koreans; (d) general practitioner referrals to specialists; and (e) the competing nature of doctoring. The interviews with the eight doctors were tape-recorded and then transcribed into a full-text report for analysis.

First, *open coding* of the data was conducted to form categories of information about the event being examined. Next, *axial coding* was performed; this step involved the researchers taking one of the categories generated during open coding and exploring it as a core phenomenon. During this phase, other categories (e.g., casual conditions, content, intervening condition, strategies, and consequences) were connected to the core phenomenon. Finally, in *selective coding*, the core category (i.e., the central phenomenon under investigation) was selected and systematically related (or integrated) with other categories. These three phases allowed for the construction of the overarching theoretical model.

Figure 11.2 Emerging Design



Source: Glaser and Strauss (1967).

Design: Qualitative method using a grounded theory perspective with a systematic design¹

Example for Figure 11.2

Zoffmann, V., & Kirkevoold, M. (2007). Relationships and their potential for change developed in difficult type 1 diabetes. *Qualitative Health Research*, 17(5), 625–638.

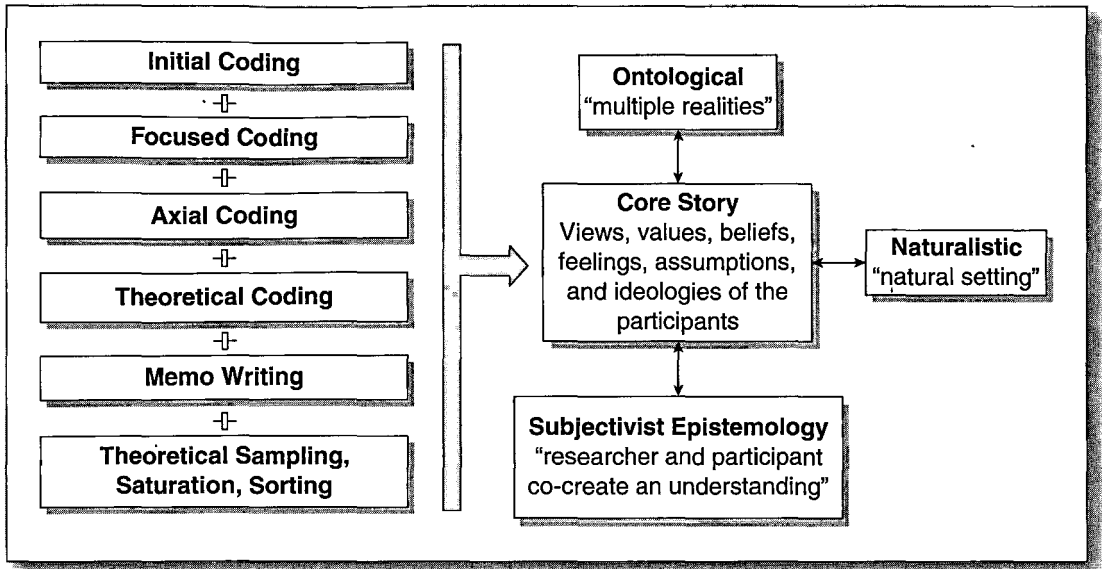
Research Aim: To develop a theory that interprets patient–provider relationships as a framework for acknowledging and exploiting the relational potential for change in difficult diabetes care.

Procedures: Dyads (one patient and one nurse) were formed based on the assignment of nurses to patients in the units. Following the principles of theoretical sampling, researchers used the first case (dyad) to generate a hypothesis. To investigate and compare the processes related to this hypothesis, subsequent patients were theoretically sampled to ensure a variation in levels of self-management resources. As the primary data sources, two patient–nurse conversations were taped from each dyad, one at the beginning and one at the end of the hospital stay. These interviews revealed the experiences, considerations, and feelings of both parties with regard to the hospital stay.

The abductive process of applying constant comparative analysis was performed. First, initial open coding of each interview was conducted. Through a combination of listening and writing, notes were created that provided ideas for the tentative advancement of more abstract codes. Second, critical comparisons were performed on the most solid categories that were supported by transcriptions of the coded data. This process was used to specify the content and further the advancement of lasting categories and subcategories. During the third step, comparisons across data sources were performed to explore and confirm links between concepts and thus pattern out theoretical connections. These initial theoretical constituents were compared in the fourth step (ongoing throughout the process of writing) to connect them to larger elements for further theory building. At each step, there was a return to former steps to test fit, work, relevance, and modifiability.

Design: Qualitative method using a grounded theory perspective with an emerging design

¹Although the authors state a “critical realist analysis” in the title, a systematic design was applied.

Figure 11.3 Constructivist Design

Source: Charmaz (2014).

Example for Figure 11.3

Jones, S. R., & Hill, K. E. (2003). Understanding patterns of commitment: Student motivation for community service involvement. *Journal of Higher Education*, 74(5), 516–539.

Guiding Research Questions: What are students' reasons for participation in community service in high school? What are students' reasons for participation in community service (or not) in college? How do students explain and understand the relationship between high school and college involvement? To what extent do students attribute differences or changes in their motivations as well as to their experiences?

Procedures: Purposive sampling was used for both the identification of college and university participants and the selection of student participants at each institution. The primary strategy for data collection was in-depth, semistructured interviewing. Questions focused on the nature of community service involvement in high school and college, the reasons students attributed to their participation, and the meaning they attached to community service. The researchers were also interested in understanding the

campus context in which students' decisions about community service participation were enacted. Thus, they examined materials from the community service programs.

All data were analyzed using the *constant comparative method*. The analytic process moves from more concrete codes to abstract themes and categories that are reflective of the meaning that participants attach to their experiences, rather than the generation of objective truth. Thus, the data analysis proceeds in a cyclical manner with the researchers constantly returning to the data with new questions and ideas until a narrative emerges that describes the essence of experience for study participants. This essence of experience is described as the core story.

Design: Qualitative method using a grounded theory perspective with a constructivist design

Reviewing the Content and Testing Your Knowledge

Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Qualitative Method** and the **Grounded Theory Perspective**. The research will be considered nonexperimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.
3. Develop the appropriate primary research question to be associated with this design.
4. Discuss the sampling strategy and technique to be used.
5. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures.
6. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
7. Briefly discuss the strengths and limitations associated with this approach and the specific design.

CHAPTER 12

ETHNOGRAPHIC PERSPECTIVE

Ethnography is an approach that was developed to describe cultures; this includes any culture that shares group characteristics such as values, beliefs, or ideas. The ethnographic researcher is interested in understanding another way of life from the point of view of the participants who make up the culture or group being studied. Because this perspective is based on understanding anything associated with human behavior and belief, it is well-suited for the fields of education and the social and behavioral sciences, including more recent areas of study like the research of culture and its relation to the Internet (see Hine, 2015).

Ethnography can be defined as research designed to describe and analyze the social life and culture of a particular social system, based on detailed observations of what people actually do. The researcher is embedded within the culture and takes a firsthand account of the beliefs, motivations, and behaviors of the individuals in the group. The data that are collected are used to (a) document the lives of the participants within the context of the culture, (b) understand the experiences of the individuals within the culture, or (c) interpret the behaviors shaped by the cultural context.

◆ REALIST DESIGN

Van Maanen (1988) stressed three aspects of the realist design: (a) the invisible author (i.e., narrating in third person), (b) thick descriptions of the

mundane (using a system of standard categories to organize the descriptions), and (c) interpretive “omnipotence” (i.e., allowing the author the final word in presenting the culture). The realist design offers one researcher’s overall perspective of a phenomenon from facts that are meticulously culled down to support a perspective. Thus, although the researcher’s duty is to objectively (without bias) present the facts, ultimately the interpretations of the facts come from the “omnipotent” researcher. In general, Spradley’s (1979, 1980) designs are less “narrative” or “literary” than those of van Maanen (1988) and Geertz (1998).

CRITICAL DESIGN ♦

The critical design allows for the critiquing (i.e., challenging the status quo) of some existing system while maintaining a level of scientific inquiry. It provides a scientific framework for advocacy or a structure for directly examining relationships among cultural features, economic systems, knowledge, society, and political action. Put simply, Madison (2011) and Thomas (1993) both asserted that the critical design is used to describe, analyze, and scrutinize hidden agendas, power centers, and assumptions that inhibit, repress, and constrain. Thus, the real utility of a critical design is the structure it provides for researchers who are interested in explaining some form of ideology or power relations through the transformation of meaning and conceptualization of existing social systems.

CASE STUDY DESIGN ♦

The case study design is often used with the ethnographic perspective; however, it has some distinct differences from traditional ethnography. While traditional ethnography is focused on group behavior, the case study design allows for the investigation of individuals as a whole (Creswell, 2012).¹ This design provides the framework for an in-depth contextual analysis of a finite number of events or conditions and their associations. More specifically, the ethnographic case study allows for the examination of an actual case within some cultural group. The “case” being explored also can be a group bound by time, place, or environment (i.e., a group

¹Creswell (2012) identified many different types of designs within the ethnographic approach, such as confessional, life history, autoethnography, microethnography, feminist, postmodern, and ethnographic novels.

must be considered a unit, which is more than just a homogenous group). Researchers interested in exploring activities of a group, rather than shared patterns of group behavior, should follow this design.

WHEN TO USE ETHNOGRAPHY

- Studying a school, organization, or program in-depth
- Studying what people do
- Studying how things work or run
- Studying “insiders”
- Studying aspects of “culture” (e.g., practices, rituals, lives, interconnections, customs, values, beliefs, everyday life)

We refer the reader to the following books for further details regarding the ethnographic approach:

Fetterman, D. M. (2009). *Ethnography: Step-by-step* (3rd ed.). Thousand Oaks, CA: Sage.

Madison, D. S. (2011). *Critical ethnography: Methods, ethics, and performance* (2nd ed.). Thousand Oaks, CA: Sage.

Makagon, D., & Neumann, M. (2008). *Recording culture*. Thousand Oaks, CA: Sage.

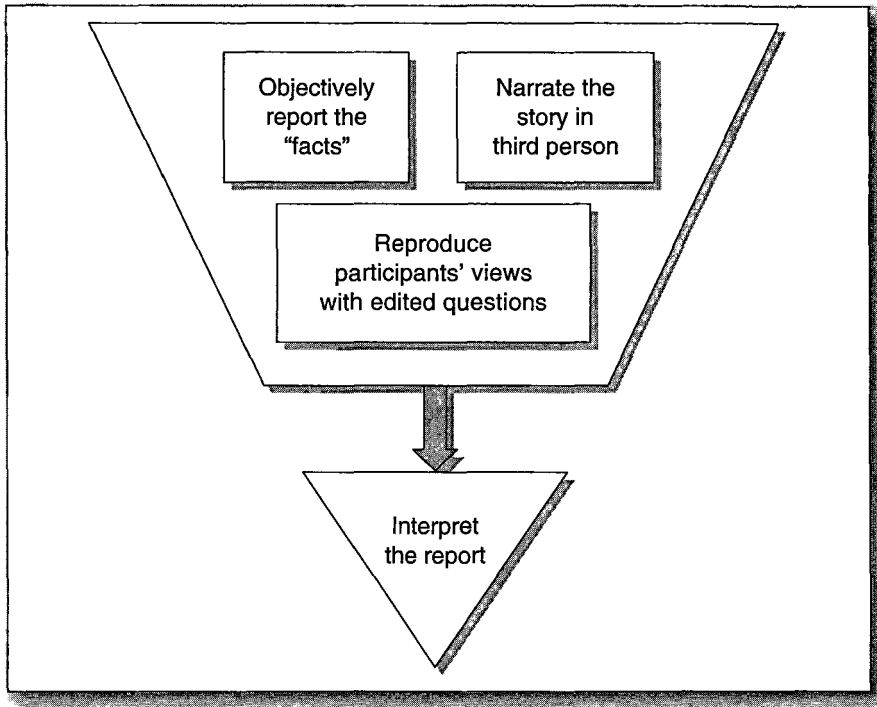
Van Maanen, J. (1988). *Tales of the field: On writing ethnography*. Chicago, IL: University of Chicago Press.

Example for Figure 12.1

Purser, G. (2009). The dignity of job-seeking men: Boundary work among immigrant day laborers. *Journal of Contemporary Ethnography*, 38(1), 117–139.

Research Aim: Examine the discourses through which Latino immigrant day laborers make sense of, and find dignity within, their ongoing quest for work.

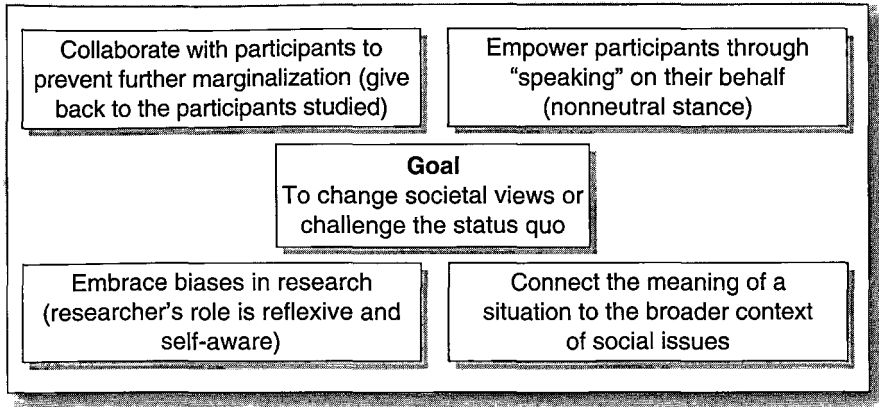
Procedures: The data collection involved ethnographic fieldwork and interviews with individual day laborers. The researcher conducted a total of 22 in-depth, loosely structured interviews with day laborers, 10 of whom regularly sought work out of the employment center and

Figure 12.1 Realist Design

12 of whom regularly sought work on the street. This was followed by a series of open-ended questions that focused on the objective and subjective dimensions of the men's work experiences and job-searching strategies. Substantial attention was devoted to understanding how the men made sense of their precarious position on the margins of the labor market.

This study involved an inductive approach to data analysis. The third-person voice was used, and no personal ideas were included in the report; rather, the facts were presented through the actual words of the participants. Objective data were reported without personal bias or judgment. Ordinary details of each participant's life were included, and standard categories for cultural description were used (e.g., work life and family systems). The final interpretive report allowed the researcher to provide subjective explanations of the data representing the culture being studied.

Design: Qualitative method using an ethnographic perspective with a realist design

Figure 12.2 Critical Design

Example for Figure 12.2

Varcoe, C., Browne, A. J., Wong, S., & Smye, V. L. (2009). Harms and benefits: Collecting ethnicity data in a clinical context. *Social Science & Medicine*, 68, 1659–1666.

Research Aim: Critically examine the implications of collecting ethnicity data in health care settings.

Procedures: Data were collected in four modes: (a) in-depth interviews with decision-makers and policy leaders affiliated with health authorities, (b) focus groups of community leaders who served on committees of the health authority to represent patients' perspectives concerning health care planning, (c) semistructured interviews with patients seeking health services in either a subacute area or a community health center, and (d) interviews with health care workers who were involved in either administering an ethnic identity question in health care agencies or whose agencies were considering doing so as part of intake data. Patient interviews were focused on their thoughts of their identification of ethnicity in health care settings, past experiences with being asked, and their thoughts on the benefits and concerns.

An interpretive thematic analysis was conducted. The theoretical perspective was guided by an ethical lens. Each transcript and associated field notes were read to get a sense of the whole and then coded thematically. Collaboration with participants occurred throughout the process to optimize the study's benefits and avoid harm. The meaning of the

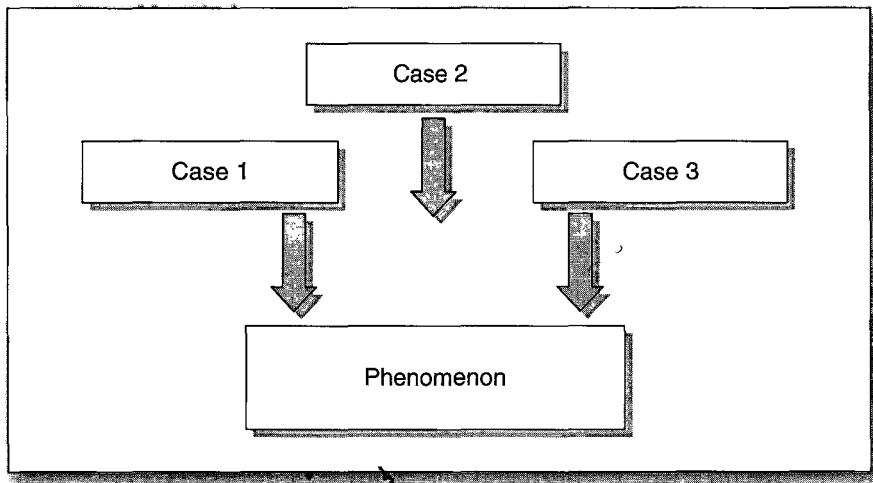
phenomenon under investigation was connected to the broader context of power and the social positions of patients within existing power structures. Themes were compared across interviews and revised, based upon the views and biases of the researcher and participants. The final report advocated on behalf of the participant group. The researchers challenged the status quo assumption that providing quality care in the clinical context requires the collection of ethnicity data. The intent was to change societal standards by a call to action and to address the structural inequities at health care settings.

Design: Qualitative method using an ethnographic perspective with a critical design

Example for Figure 12.3²

Richardson, J. B. (2009). Men do matter: Ethnographic insights on the socially supportive role of the African American uncle in the lives of inner-city African American male youth. *Journal of Family Issues*, 30(8), 1041–1069.

Figure 12.3 Case Study Design



²According to Creswell (2012), this example is a collective case study, with several (there can be more than three) cases that provide insight into a phenomenon. There is also an (a) intrinsic case study design, which explores a single unusual case, and (b) an instrumental case study design, which examines a single case to gain insight into a phenomenon.

Research Question: What is the role of the African American uncle as a form of social support and social capital in the lives of adolescent African American males living in single-female-headed households?

Procedures: The study involved in-depth life-history interviews and ethnographic participant observations of young men and their single mothers over a period of 4 years. The use of ethnography provided exhaustive and rich contextual data. The qualitative inquiry highlights the contextual nature of social life; it explores subjective perceptions and meanings, and it identifies social processes and dynamics. In the three cases studied, the young men in the sample and their mothers were able to clearly identify and explain the socially supportive role that uncles filled as surrogate fathers. In some instances, the uncles themselves articulated their roles as surrogate fathers.

First, a phenomenon (role of the African American uncle as a form of social support) was identified. Next, the appropriate case(s) (individual, activity, event, or process) were chosen that allowed for the examination of the phenomenon. The description and comparisons of three cases helped to provide insight into the role of the African American uncle as a form of social support in the lives of adolescent African American males (i.e., the phenomenon). Multiple forms of data were collected to increase the depth of understanding regarding the phenomenon of interest. The cases were also presented within a larger context (setting, political climate, social and economic status).

Design: Qualitative method using an ethnographic perspective with a case study design

Reviewing the Content and Testing Your Knowledge

Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Qualitative Method** and the **Ethnographic Perspective**. The research will be considered nonexperimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.

3. Develop the appropriate primary research question to be associated with this design.
4. Discuss the sampling strategy and technique to be used.
5. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures.
6. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
7. Briefly discuss the strengths and limitations associated with this approach and the specific design.

CHAPTER 13

NARRATIVE PERSPECTIVE

The narrative approach involves gathering information, in the form of storytelling by the participant, for the purpose of understanding a phenomenon. Humans are storytelling beings by nature; we lead storied lives, both individually and collectively. Ultimately, the narrative approach is most widely used in the disciplines of psychology and psychiatry and is the study of the multitude of ways humans experience the world. Specifically, this approach involves collaboration between the researcher and participant, as a way to understand phenomena through stories lived and told. The narrative design involves (a) the exploration of a single participant or a small sample of participants, (b) gathering data through the collection of stories, (c) retelling the stories (restorying), and (d) reviewing the story with the participant to help validate the meaning and subsequent interpretation. The narrative design can be either biographical or autobiographical.

Dialogic listening skills are essential to the narrative approach; this type of “listening” is used throughout the whole process, as the researcher gathers data through conversations and engaged interchanges of ideas and information with the participant(s). The narrative approach can be conceptualized as descriptive, explanatory, or critical by design and follows the “underlying assumptions that there is neither a single, absolute truth in human reality nor one correct reading or interpretation of a text” (Polkinghorne, 1988, p. 2). There is also a structural approach in the way individual stories are studied (Riessman, 2007).

Dan McAdams (creating self in narrative) and Jefferson Singer (explanatory potential of the life story) have had a profound influence on the development and use of the narrative approach within the social and behavioral sciences. McAdams, Josselson, and Lieblich's (2006) contributions included (a) the Life Story Interview method, (b) the Guided Autobiography, (c) the Loyola Generativity Scale, and (d) a set of coding manuals to analyze the stories of research participants. Singer's (1997) book *Message in a Bottle* focused on men whose addictions were resistant to the traditional 12-step method and served as an excellent exemplar of the narrative approach. Singer also used the explanatory potential of the life story of individuals within the therapeutic context (see Singer & Bonalume, 2010a and 2010b, for more on autobiographical narrative approaches for case studies in psychotherapy).

DESCRIPTIVE DESIGN ♦

The descriptive design involves the description of any one or more of the following: (a) individual or group narratives of life stories or specific life events, (b) the conditions or contextual factors supporting the story, (c) the relationship between individual stories and the culture the stories are embedded within, and (d) how certain life events impact the participants' story line. Thus, the descriptive design is used to explore the status of some phenomenon and to describe what exists with respect to the individual, group, or condition.

EXPLANATORY DESIGN ♦

The explanatory design is used to provide an account of some phenomenon by means of why something happened. Thus, the explanatory design is used to explore the causes and reasons of phenomena.

CRITICAL DESIGN ♦

Van Maanen (1988), in his book on ethnography, discussed the use of "critical tales." These critical tales are conceptualized as narrative approaches using a critical framework. A critical tale may illuminate individual experiences as

well as larger social, political, symbolic, or economic issues. Thus, the critical design within the narrative approach involves the same structure or framework as the critical design within the ethnographic approach. Ultimately, this design allows for the critiquing of some existing system while maintaining a level of scientific inquiry.

WHEN TO USE NARRATIVE INQUIRY

- Telling stories about stories
- Exploring identity and conflict
- Examining the structure of experience
- Focusing on how people create meaning in their lives
- Exploring the interaction of individual stories with cultural narratives

We refer the reader to the following books for further details regarding the narrative approach:

Clandinin, D. J., & Connelly, F. M. (2004). *Narrative inquiry: Experience and story in qualitative research*. San Francisco, CA: Wiley.

Lieblich, A., Tuval-Mashiach, R., & Zilber, T. (1998). *Narrative research: Reading, analysis, and interpretation*. Thousand Oaks, CA: Sage.

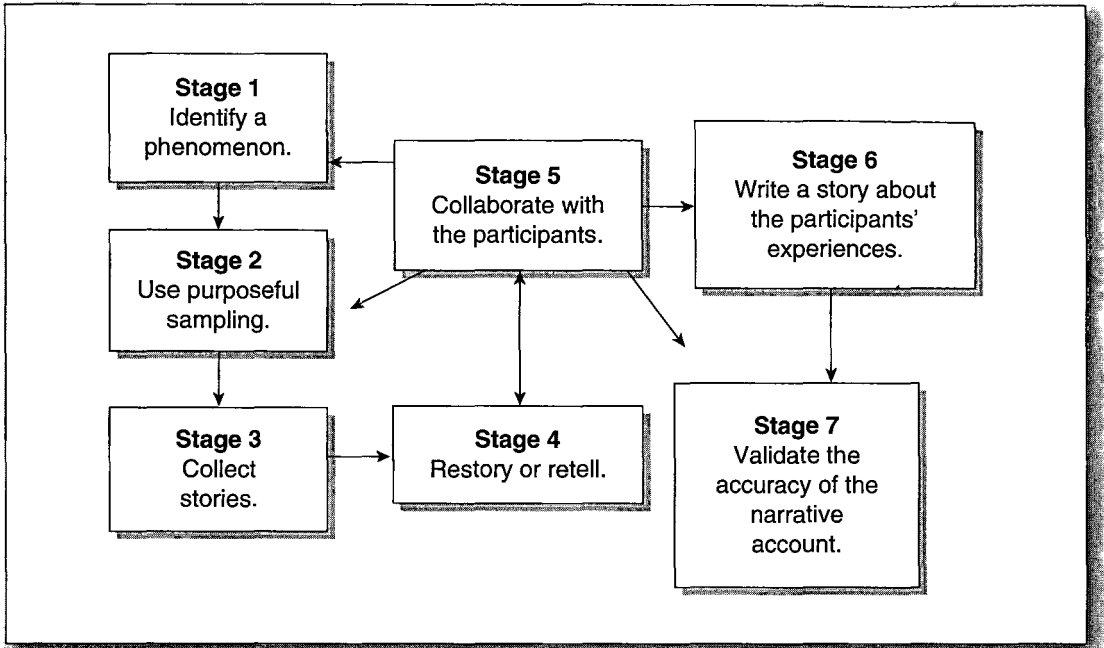
Riessman, C. K. (2007). *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage.

Example for Figure 13.1

Lapadat, J. C. (2004). Autobiographical memories of early language and literacy development. *Narrative Inquiry*, 14(1), 113–140.

Research Aim: To explore adults' memories of their own acquisition of language and literacy learning

Procedures: Participants kept a journal in which they made regular entries over the semester, reflecting on their own personal history of learning language and literacy from the preschool years through the end of adolescence. The participants were asked to recall personally significant events, situations, and people that made a difference to their learning, as well as

Figure 13.1 Descriptive Design

Note: At Stage 5 of the process, the researcher can revert to any previous stage as indicated by the arrows.

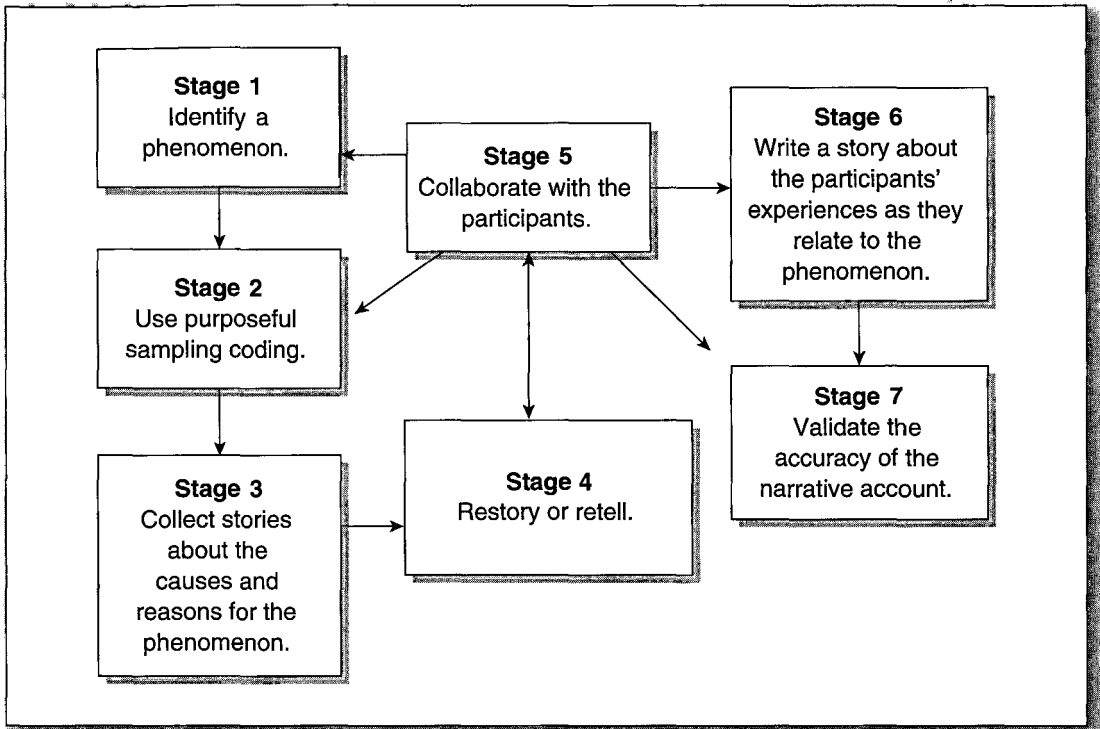
ways in which their learning and use of language made a difference in their lives. The participants were asked to structure entries around particular topics they set for themselves, to avoid holding tightly to a chronological sequence, and to discuss specific examples.

In Stage 1, the researchers chose to explore the acquisition of early language and literacy. Purposeful sampling (Stage 2) was conducted (i.e., adults in a language development seminar). During Stage 3, stories from the participants were collected (e.g., personal history of learning language and literacy). The fourth stage involved the identification of categories (e.g., family and home, peers and friends, school and teachers, books and becoming literate, culture and languages) and then *restorying* by sequencing and organizing the elements of the story identified by the researcher (e.g., poetic transcription). The fifth stage occurred throughout the process, as the researcher collaborated with the participant to ensure the validity of the individual experiences. Stage 6 involved the use of the first person to complete the narrative report. During Stage 7, the researcher

consulted with the participants to ensure the accuracy of the final narrative account.

Design: Qualitative method using a narrative perspective with a descriptive design

Figure 13.2 Explanatory Design



Example for Figure 13.2

Hirakata, P. (2009). Narratives of dissociation: Insights into the treatment of dissociation in individuals who were sexually abused as children. *Journal of Trauma & Dissociation*, 10(1), 297–314.

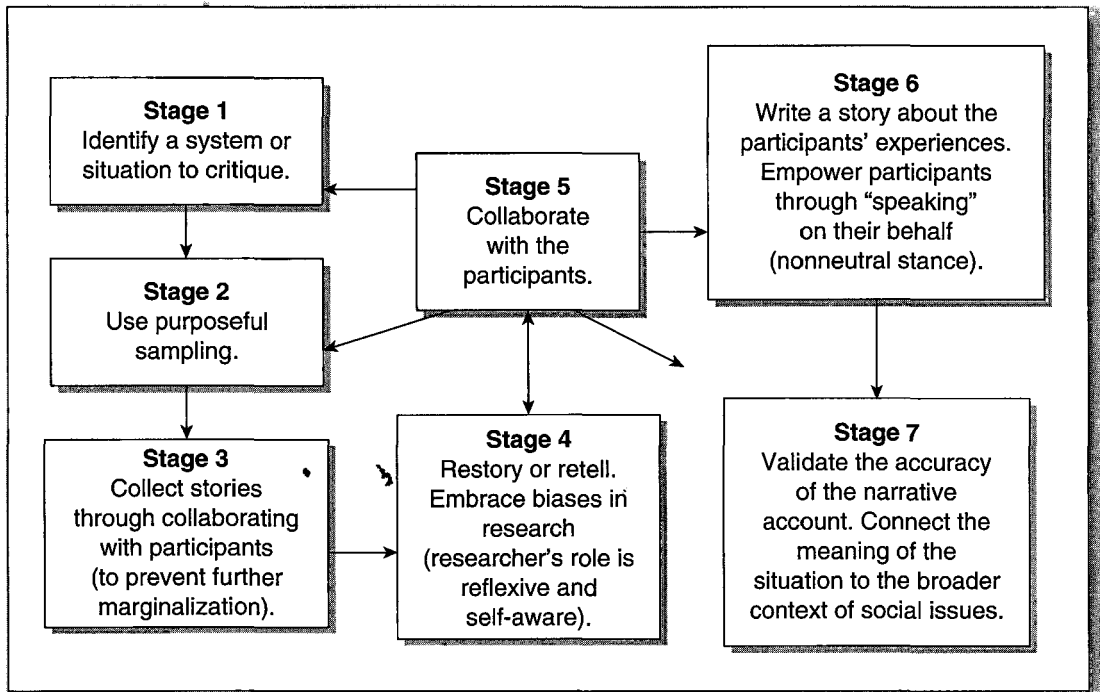
Research Aim: To explore the treatment of dissociation and provide support for treatment approaches that are viewed as helpful by clients who engage in dissociative behavior

Procedures: Participants engaged in a single interview (cross-sectional) designed to identify factors that positively or negatively influenced

therapy and minimized the overall need to dissociate. Interviews were semistructured and included questions such as “How was dissociation addressed in therapy?”, “What did you find helpful?”, and “What did you find not helpful or even harmful?” The interviews were also audiotaped and transcribed, and the data were analyzed using a holistic-content approach (Lieblich, Tuval-Mashiach, & Zilber, 1998): Each interview was read for its content in a holistic manner until patterns, or themes, began to emerge. Global impressions were noted by the space dedicated to a certain issue and the repetitive nature that occurred both within and across narratives. Any exceptions or unusual and contradictory features were also recognized. Each participant read and validated his or her individual narrative.

Three major themes emerged from this study: (a) tools and techniques, (b) a nonpathologizing approach, and (c) the therapeutic relationship. These themes were further divided into 16 subthemes. Each subtheme was discussed individually, and excerpts from the participants’ interviews were provided. The findings from this study were explanatory in nature, providing insight into the treatment of dissociation.

Figure 13.3 Critical Design



Design: Qualitative method using a narrative perspective with an explanatory design

Example for Figure 13.3

Berman, H., Mulcachy, G. A., Forchuk, C., Edmunds, K. A., Haldenby, A., & Lopez, R. (2009). Uprooted and displaced: A critical narrative study of homeless, aboriginal, and newcomer girls in Canada. *Issues in Mental Health Nursing, 30*(7), 418–430.

Research Aim: Examine how uprooting and displacement have shaped mental health among three groups: (a) newcomers to Canada (immigrant and refugee girls), (b) homeless girls, and (c) Aboriginal girls.

Procedures: During face-to-face dialogic interviews, the researchers explored the means and strategies of uprooting and displacement and how these experiences affected the participants. Dialogic and reflective techniques were used to allow the respondents to become actively involved in the construction and validation of meaning (Maguire, 1987). An interview guide was used flexibly, with probes to encourage dialogue, critical reflection, and elaboration of responses. The research team was there to establish the context for the interview, offering overall direction and providing affirming feedback. However, the open-ended structure to the narrative interview allowed the participants to direct the flow and focus of the conversation.

All participants were given the choice of being interviewed alone or in a small group consisting of two to four girls. The researchers' rationale for this option was the potential power of group interviews to provide a context in which individuals are able to analyze the struggles they had encountered and the challenges they had faced, to simultaneously begin to collectivize their experiences, and to develop a sense of empowerment as they began to see the possibilities for change. The researchers analyzed the participants' words systematically, in line with the narrative approach.

The researchers were guided by the supposition that the stories told by participants would provide insight into the perceptions about what happened to them, as well as the social, economic, and political meanings of those events. Consistent with the tenets of a critical design, the researchers' goal was to facilitate the development of knowledge in ways that had the potential for emancipation and empowerment.

Design: Qualitative method using a narrative perspective with a critical design

Reviewing the Content and Testing Your Knowledge

Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Qualitative Method** and the **Narrative Perspective**. The research will be considered nonexperimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.
3. Develop the appropriate primary research question to be associated with this design.
4. Discuss the sampling strategy and technique to be used.
5. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures.
6. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
7. Briefly discuss the strengths and limitations associated with this approach and the specific design.

CHAPTER 14

PHENOMENOLOGICAL PERSPECTIVE

Phenomenology, put simply, is the description of an individual's immediate experience. The phenomenological approach was born out of Edmond Husserl's philosophical position that the starting point for knowledge was the self's experience of phenomena, such as one's conscious perceptions and sensations that arise from life experience. From this philosophy emerged the modern-day phenomenological approach to research with the goal of understanding how individuals construct reality. Researchers use the phenomenological approach when they are interested in exploring the meaning, composition, and core of the lived experience of specific phenomena. The researcher explores the conscious experiences of an individual in an attempt to distill these experiences or get at their essence.

◆ EXISTENTIAL DESIGN

The aim is to illuminate the essential general meaning structure of a specific phenomenon, with a focus on grasping the whole meaning of the experience, instead of dividing it into parts. Researchers using the existential design move from the concrete description of the experience of a given participant (co-researcher) to the interpretation of said experience. The participants (co-researchers) are asked for a description of their concrete experiences. The ultimate goal is to comprehend human experience as it is

actually lived in the “real world” rather than in some artificial environment (von Eckartsberg, 1997).

Basic themes of existential phenomenology are (a) lived experience, (b) modes of being, and (c) ontology (the study of the nature of being, existence, or reality). In fact, the existential phenomenology associated with Heidegger’s philosophy is often referred to as *ontological phenomenology*, as it is primarily concerned with “being.” This differs from transcendental phenomenology, which is most associated with Husserl’s epistemological philosophy (concerned with knowledge).

TRANSCENDENTAL DESIGN ♦

Some key tenets of the transcendental design are (a) intentionality (consciousness is always intentional), (b) eidetic reduction (researcher accesses the consciousness of the participant to get at the pure essence of some phenomenon, thus revealing the essential structure), and (c) constitution of meaning (returning to the world from consciousness). This design is descriptive in nature, as it is through analysis and description of how things are constituted in, and by, consciousness that allows us to understand various phenomena. This design is useful for researchers who are interested in gathering data to grasp the essence of the human experience.

HERMENEUTIC DESIGN ♦

Some key tenets of the hermeneutic design are (a) interpretation, (b) textual meaning, (c) dialogue, (d) pre-understanding, and (e) tradition. The hermeneutic design deviates from the descriptive nature of which the phenomenological approach is most often associated. This design has a strong focus on reflective interpretation, made evident by Heidegger, who asserted that description is inextricably linked to interpretation. Essentially, this design is based on the fundamental theory that all forms of human awareness are interpretive.

CASE STUDY DESIGN ♦

The case study design is also often used with the phenomenological perspective. This design lends itself well to the exploration of meaning of

a lived experience of some phenomenon. This design provides the framework for an in-depth analysis of a finite number of participants. Researchers interested in exploring activities of an individual or small group, rather than the shared patterns of group behavior, should follow this design.

WHEN TO USE PHENOMENOLOGY

- Studying people's experiences
- Studying how people make meaning in their lives
- Studying relationships between what happened and how people have come to understand these events
- Exploring how people experience the essence of a particular phenomenon
- Examining the commonalities across individuals

We refer the reader to the following books for further details regarding the phenomenological approach:

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretive phenomenological analysis: Theory, method, and research*. Thousand Oaks, CA: Sage.

Vagle, M. D. (2014). *Crafting phenomenological research*. Walnut Creek, CA: Left Coast Press.

van Manen, M. (2014). *Phenomenology of practice: Meaning-giving methods in phenomenological research and writing*. Walnut Creek, CA: Left Coast Press.

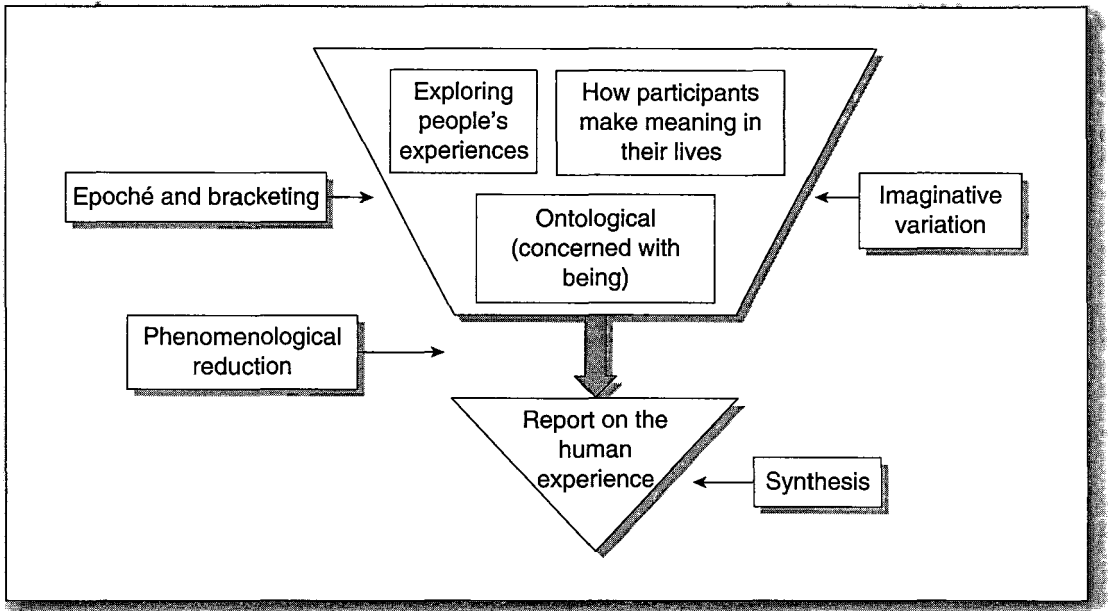
Example for Figure 14.1

Smith, M. E. (2007). Self-deception among men who are mandated to attend a batterer intervention program. *Perspectives in Psychiatric Care*, 43(4), 193–203.

Research Aim: Gain an understanding of the perceptions of perpetrators of intimate partner violence (IPV) prior to beginning a Batterers' Intervention Program (BIP).

Procedures: Qualitative methods used in this study were conducted according to the existential-phenomenological method outlined by Pollio, Henley, and Thompson (1997). The method of existential phenomenology was used in this study to provide men the opportunity to describe their perceptions concerning the meaning attached to being mandated to attend a BIP.

Figure 14.1 Existential Design

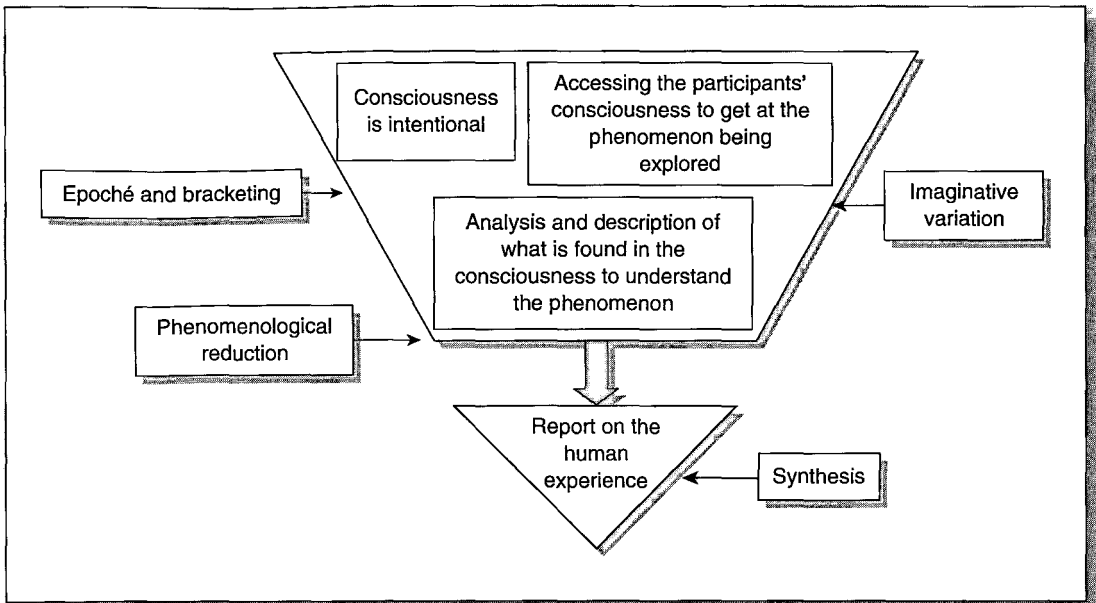


Participants were asked to fill out a demographic questionnaire and then participate in a face-to-face, audiotaped interview. Prior to beginning the study, the first author of Pollio et al. (1997) participated in an individual bracketing interview in order to become more aware of her own biases as a result of her clinical practice with women who experience IPV. Men participated in a face-to-face interview after being mandated to attend a BIP but before attending their first class.

The interviews began with the prompt "Tell me about your experiences that brought you to a batterers' intervention program." Except for the initial question, all questions flowed from the dialogue. Additional questions were limited to areas of clarification and/or elaboration.

The respondents' own words were used to support a given interpretation. Participant transcripts were then related to each other to identify common patterns or global themes. All of the transcripts were read in a group context to reduce researcher bias. Subsequently, an overall thematic description was developed of the meaning the perpetrator attached to attending a BIP.

Design: Qualitative method using a phenomenological perspective with an existential design

Figure 14.2 Transcendental Design

Example for Figure 14.2

Moerer-Urdahl, T., & Creswell, J. W. (2004). Using transcendental phenomenology to explore the “ripple effect” in a leadership mentoring program. *International Journal of Qualitative Methods*, 3(2), 19–35.

Research Aim: To understand the meaning of mentors’ experiences with the ripple effect and their experiences of reinvesting in others.

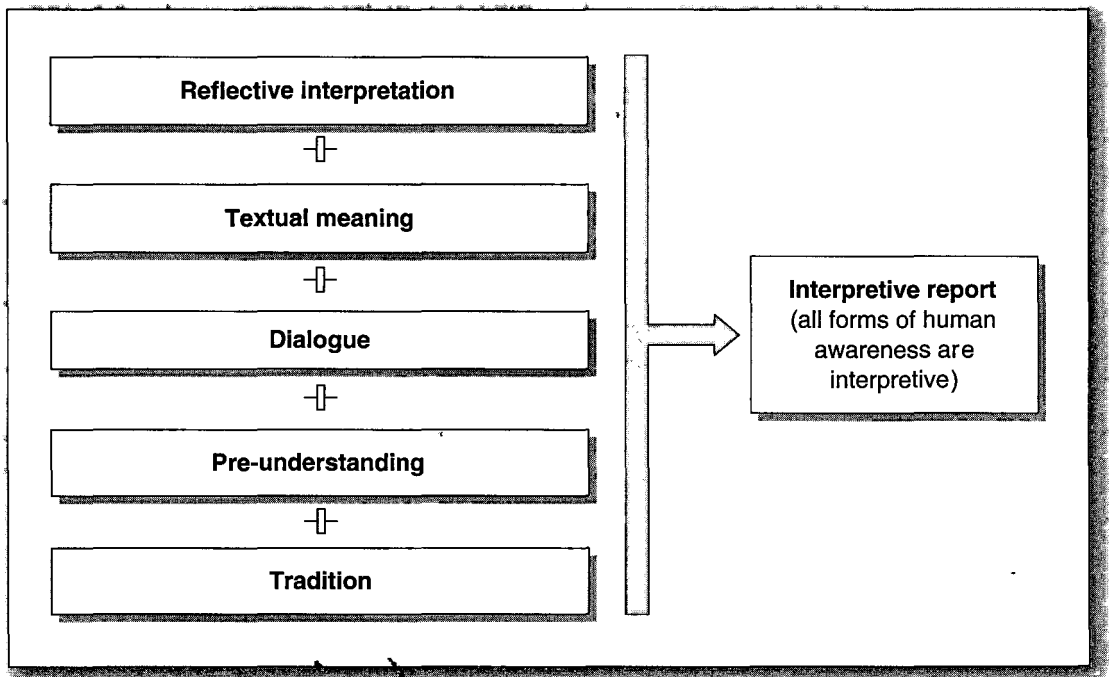
Procedures: Two central questions in this study address key questions that van Manen (2014) recommended that phenomenologists ask: What were their experiences with the ripple effect? And in what context or situations did they experience it? They were also asked if they considered themselves to be mentors today and, if so, to whom, and if they had been mentored in the past, and by whom.

Detailed telephone interviews were conducted with nine participants, and these interviews were audiotaped, lasting for 25 to 50 minutes. Transcendental phenomenology was chosen as the appropriate methodology for this research as the researchers were searching for an understanding of the meaning of the participants’ experiences. Additionally, the systemic procedures and detailed data analysis steps as outlined by Vagle (2014) are

ideal for assisting less experienced researchers. The researchers set aside prejudgments in a process called *epoché*, a Greek word meaning to refrain from judgment. The researchers described their own experiences with the phenomenon, identified significant statements in the database from participants, and clustered these statements into meaning units and themes (*epoché* and bracketing). Next, the researchers synthesized the themes into a description of the experiences of the individuals (textual and structural descriptions) and then constructed a composite description of the meanings and the essences of the experiences.

Design: Qualitative method using a phenomenological perspective with a transcendental design

Figure 14.3 Hermeneutic Design



Example for Figure 14.3

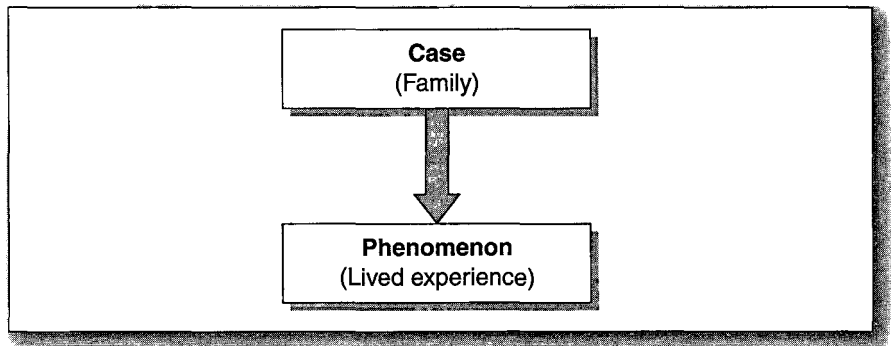
Shin, K. R. (2002). Using hermeneutic phenomenology to elicit a sense of body at mid-life. *International Journal of Qualitative Methods*, 1(2), 39–50.

Research Aim: To seek the essence of women's experience of bodily changes caused by menopause and thereby to provide a conceptual framework for women's health promotion education program and theory development

Procedures: This study was a phenomenological and hermeneutic perspective of Korean women who experienced menopause. After exploring the literature, the researcher reflected on artistic depictions of midlife women and conversations with menopausal women in Korea. Phenomenological conversations were conducted with six menopausal women, including two in-depth, tape-recorded interviews and observations. During the initial meeting with participants, the researcher explained the purpose of the study, and then interviewees gave their informed consent to participate in the study. All participants were guaranteed anonymity. Van Manen's (1990) thematic analysis and line-by-line approach, by which every statement of the participants is thoroughly examined, was used to find what their words or sentences implied about their experiences.

Design: Qualitative method using a phenomenological perspective with a hermeneutic design

Figure 14.4 Case Study Design



Example for Figure 14.4¹

Ryndak, D. L., Storch, J. F., & Hoppey, D. (2008). One family's perspective of their experiences with school and district personnel over time related to

¹According to Creswell (2012), this is an example of an intrinsic case study design, which explores a single unusual case. There is also an instrumental case study design that examines a single case to gain insight into a phenomenon and a collective case study that uses several cases to provide insight into a phenomenon.

inclusive educational services for a family member with significant disabilities. *International Journal of Whole Schooling*, 4(2), 29–51.

Research Aim: To understand the impact of placement and special education services, especially in relation to receiving services in inclusive general education classes

Procedures: The study used a phenomenological lens to explore the experience of one family (mother, father, son, and a daughter with significant disabilities) as they sought, lived through, and reflected upon placement and services for the daughter with disabilities. The study focused on how the family understood, developed, and socially constructed meanings from the events and interactions that occurred over time as one of the family members moved from receiving special education services and supports in a more restrictive context to receiving special education services and supports in more inclusive contexts.

Phenomenological interviews were conducted three times with each of the daughter's family members (mother, father, and brother). The interviews included open-ended questions to build upon and explore each participant's past and present. Interviews were scheduled 2 weeks apart for each family member so that the researchers had time to inquire in-depth into the family's lived experiences, while at the same time providing space for the family members to share their own unique insights, stories, and experiences. Thus, each participant was able to reconstruct his or her own experience over time and construct meanings of their own experiences. The researchers then viewed these individual experiences collectively in an attempt to understand the meaning and essence of the family member's collective experiences.

After each interview, the researcher shared the transcript with the interviewee and requested that the family member check the transcript for accuracy, making additions and deletions to further clarify his or her experiences and perspectives. The researchers used Wolcott's (1994) approach of description, analysis, and interpretation as a method for making sense of interview data. The constant comparison method of reflecting and exploring the data allowed emerging patterns to collectively come into focus.

Design: Qualitative method using a phenomenological perspective with a case study design

Reviewing the Content and Testing Your Knowledge

Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Qualitative Method** and the **Phenomenological Perspective**. The research will be considered nonexperimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.
3. Develop the appropriate primary research question to be associated with this design.
4. Discuss the sampling strategy and technique to be used.
5. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures.
6. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
7. Briefly discuss the strengths and limitations associated with this approach and the specific design.

PART IV

Mixed Methods

This part includes four popular approaches to mixed methods, followed by some of the associated basic designs (accompanied by brief descriptions of published studies that use the design). Most of the diagrams of mixed method designs were compiled and adapted from two major sources (Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010a). Visit the companion website at study.sagepub.com/edmonds2e to

<i>Method</i>	<i>Mixed</i>				
▼	▼				
Research	Experimental	Quasi-Experimental		Nonexperimental	
▼	▼				
Approach	Convergence-Parallel	Embedded	Explanatory-Sequential	Exploratory-Sequential	
▼	▼				
Design	Parallel-Databases	Data-Transformation	Data-Validation	Multilevel	Embedded-Experiment
	Embedded-Correlational	Embedded Case Study		Follow-Up	Explanations
	Participant-Selection	Theory-Development		Instrument-Development	Treatment-Development

Note: Mixed methods for experimental, quasi-experimental, and nonexperimental research are shown here, followed by the approach and then the design.

access valuable instructor and student resources. These resources include PowerPoint slides, discussion questions, class activities, SAGE journal articles, web resources, and online data sets.

Mixed methods examinations combine various aspects of quantitative and qualitative methods (often referred to as quantitative and qualitative *strands*). Because this type of methodology mixes both the quantitative and qualitative method, the logic of inquiry may include the use of induction, deduction, and abduction. This approach allows researchers to further examine constructs at a “deeper” level, where the quantitative strand reveals what the qualitative strand leaves out and vice versa. From a philosophical viewpoint, a link has been established between pragmatism and mixed methods. That is, this method was developed as an attempt to legitimize the use of multiple methodological strategies when answering research questions within a single study, which is considered a more practical approach to research. To conduct a sound mixed method study, it is critical that the researcher has a firm understanding of the distinguishing characteristics of quantitative methods (deduction, confirmation, hypothesis testing, explanation, prediction, standardized data collection, and statistical analysis) and qualitative methods (induction, discovery, exploration, theory generation, *researcher as an instrument* of data collection, and qualitative analysis). The mixed method includes the collection and analyses of quantitative (closed-ended and numerical) and qualitative (open-ended and textual) data (i.e., a quantitative and qualitative research question must be posed, individually analyzed and interpreted, and followed up with an overall interpretation).

Creswell and Plano Clark (2011) noted that one of the primary objectives in designing a mixed method study is to determine if the design should be fixed or emergent. Specifically, a fixed mixed method design is applied when the researcher predetermines the application and integration of a qualitative and quantitative method within a study. On the other hand, an emergent design is conducted when a researcher decides to include a qualitative or quantitative strand within an ongoing examination, purely based on necessity. The presentation within this guide falls more toward a typology-based approach—that is, we emphasize the various designs that are classified and developed for the use and application for mixed methods studies. Following those tenets, a researcher must consider many aspects regarding the implementation of a mixed method study. For example, the priority (or emphasis) of the qualitative and quantitative strands should be considered. Equal emphasis can be placed on the qualitative and quantitative method, or one strand can take priority over the other. The timing of the strands is also relevant. The strands can be implemented concurrently, sequentially,

nested (embedded), or multilayered. The mixing of the strands should also be considered, which can happen during the interpretation phase, data collection, data analysis, or at the level of the design.

Philosophically speaking, although mixed methods studies are considered pragmatic, researchers should still be cautious when using the typology-based approach to mixed method research (Collins & O’Cathain, 2009). Many of the “established” mixed methods designs may not fully address the needs of the variety of research scenarios across the various disciplines. Therefore, a more general or generic approach may be warranted. Tashakkori and Teddlie (2010a) produced a sound matrix of the various mixed methods designs, which is summarized in Appendix C. However, most of the designs presented within this part follow Creswell and Plano Clark’s (2011) design typology. Within their book, they referred to these designs as *variants*. In addition, in an earlier version of their text, they referred to these designs as *models*, many with different names than those seen in their more recent text.

MIXED METHODS LEGEND ♦

The following notations are used in the depiction of the various mixed methods approaches and designs:

<i>Design Notation</i>	<i>Design Element</i>
QUAN	Quantitatively driven study
QUAL	Qualitatively driven study
quan	Quantitative data is secondary to qualitative data
qual	Qualitative data is secondary to quantitative data
+	Indicates that quantitative and qualitative data are collected concurrently
→	Indicates that quantitative and qualitative data are collected sequentially
()	Parentheses indicate that one method is embedded within an emphasized method such as QUAL(quan)
→←	Indicates that the methods are implemented recursively
[]	Brackets indicate a mixed method study that is within a series of studies
=	Indicates the transition to mixing methods

Note: The mixed method legend is based on a notation system developed by Morse (1991), Plano Clark (2005), Nastasi et al. (2007), and Morse and Niehaus (2009).

We refer the reader to the following books and article for further details regarding mixed methods:

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

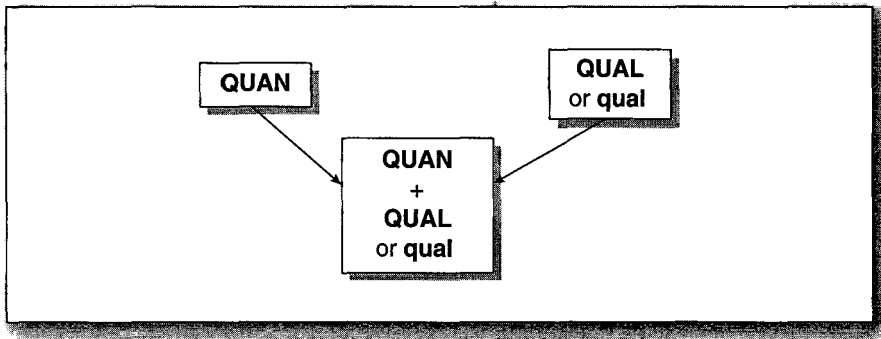
Morse, J. M., & Niehaus, L. (2009). *Mixed methods design: Principles and procedures*. Walnut Creek, CA: Left Coast Press.

Palinkas, L. A., Horwitz, S. A., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2014). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administrative Policy and Mental Health*. doi:10.1007/s10488-013-0528-y

Tashakkori, A., & Teddlie, C. B. (2010a). *SAGE handbook of mixed methods in social & behavioral research* (2nd ed.). Thousand Oaks, CA: Sage.

CHAPTER 15

CONVERGENT-PARALLEL APPROACH



The convergent-parallel approach is a *concurrent* approach and involves the simultaneous collection of qualitative and quantitative data (usually both QUAL and QUAN are the emphasis), followed by the combination and comparisons of these multiple data sources (i.e., the two methods are ultimately merged). This approach involves the collection of different but complementary data on the same phenomena. Thus, it is used for the converging and subsequent interpretation of quantitative and qualitative data. This approach is often referred to as the *concurrent triangulation design* (single-phase) because the data is collected and analyzed individually but at the same time.

♦ PARALLEL-DATABASES DESIGN

The parallel-databases design is structured so the QUAN and QUAL data are collected separately (not within the same measures) but at the same time (concurrently). The analyses of data are also analyzed concurrently. The results are then converged by comparing and contrasting the data en route to one overall interpretive framework. This design allows researchers to validate data by converging the QUAN results with the QUAL findings. This design is also referred to as a *triangulation design* and *convergence model*, as seen in Seifert, Goodman, King, and Baxter Magolda's (2010) examination.

♦ DATA-TRANSFORMATION DESIGN

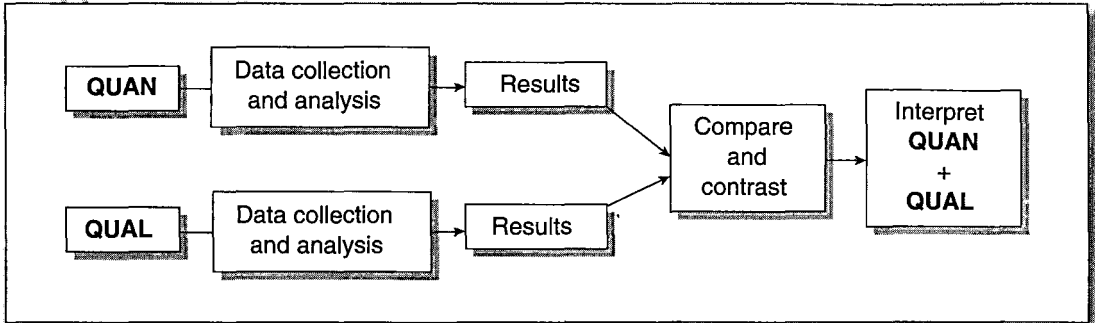
The data-transformation design allows the researcher to collect QUAN and QUAL data separately but concurrently. Following the subsequent analyses, data are transformed by either transforming QUAN to QUAL or QUAL to QUAN. Therefore, the data are mixed during this stage, followed by the subsequent analyses.

♦ DATA-VALIDATION DESIGN

The validating quantitative data design is used to validate QUAN data with qual findings. Data from QUAN and qual are collected together (within the same measures), not separately. Within this design, the qual findings are not the emphasis; therefore, they are not subject to rigorous data reduction or analysis.

♦ MULTILEVEL DESIGN

The multilevel design was originally introduced by Tashakkori and Teddlie (2002). This design allows the researcher to use different methodological techniques for addressing QUAN and QUAL data within a system. The QUAN results and QUAL findings from each level are then merged to provide an overall interpretation.

Figure 15.1 Parallel-Databases Design

Note: Any research design designated as experimental, quasi-experimental, or nonexperimental research can be used for the QUAN phase, and any approach designated under the qualitative method can be used for the QUAL phase.

Example for Figure 15.1

Hall-Kenyon, K. M., Bingham, G. E., & Korth, B. B. (2009). How do linguistically diverse students fare in full- and half-day kindergarten? Examining academic achievement, instructional quality, and attendance. *Early Education and Development, 20*, 25–52.

Research Questions

Quantitative: What are the effects of instructional quality on academic achievement (math and literacy)?

Qualitative: What are teacher and administrator perceptions of full- and half-day programs, and how do teachers' instructional behaviors influence academic achievement?

Mixed Method: To what extent do the QUAN data and QUAL data converge?

Procedures

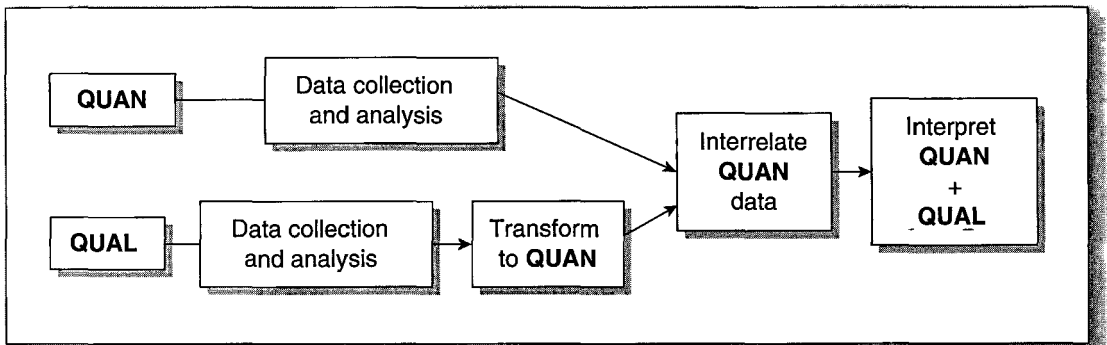
Quantitative: Eight kindergarten classrooms were used—four full-day classrooms and four half-day classrooms. The four kindergarten classrooms from the treatment school were in the first year of implementing a full-day pilot program. The four classrooms were receiving a half-day program as usual. Students' academic achievement was assessed prior to and at the

end of the school year using the following measures: (a) Peabody Picture Vocabulary Test, (b) Phonological Awareness Literacy Screening, and (c) the Applied Problems subtest of the Woodcock-Johnson III.

Qualitative: To assess the instructional quality of the teaching programs, the Classroom Assessment Scoring System was used, which includes observing teachers' instructional behaviors. Additionally, teachers and administrators from both schools were interviewed at the end of the school year using a structured interview protocol. Interview transcriptions were analyzed inductively by two of the three authors. The two authors independently read all of the interview transcriptions and coded each separate idea generated by the teachers and administrators into categories.

Design: A mixed methods study using nonexperimental research with a convergent-parallel approach and a parallel-databases design¹

Figure 15.2 Data-Transformation Design



Note: Any research design designated as experimental, quasi-experimental, or nonexperimental research can be used for the QUAN phase, and any approach designated under the qualitative method can be used for the QUAL phase.

Example for Figure 15.2

Kaldjian, L. C., Jones, E. W., Rosenthal, G. E., Tripp-Reimer, T., & Hillis, S. L. (2006). An empirically derived taxonomy of factors affecting physicians' willingness to disclose medical errors. *Journal of General Internal Medicine, 21*, 942–948.

¹The authors refer to the design as a concurrent triangulation strategy, which follows the parallel-databases design.

Research Questions

y

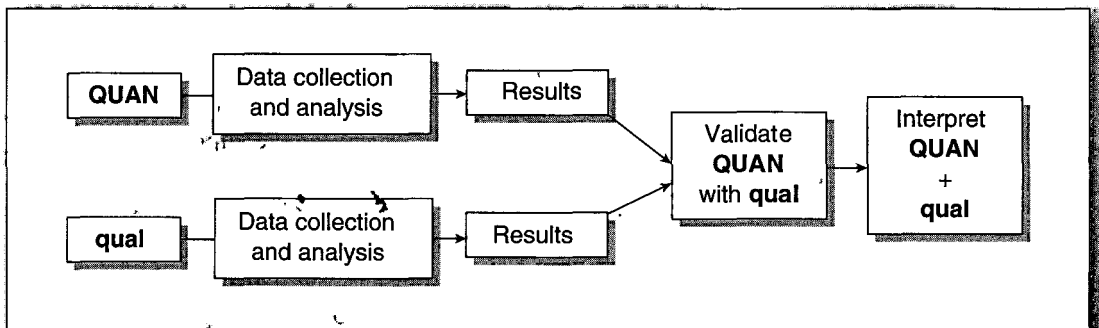
Quantitative and Qualitative: What is the taxonomy of factors that affect voluntary disclosure of errors by physicians?

Mixed Method: To what extent do the same types of data (QUAN and QUAL) confirm each other?

Procedures: Initially, a literature review was conducted. Articles were selected that addressed the experience of selected physicians. Specific factors were identified and selected from the review process. Labels for facilitating and impeding factors were derived through an iterative process. The iterative process of naming served to synthesize the linguistic heterogeneity of the literature reviewed. Next, focus groups were conducted to discuss factors related to physician self-disclosure of medical errors to institutions, patients, and colleagues. The factors identified from the focus groups were combined with the factors from the literature review. The next step included a pile-sorting task. The pile-sorting results were entered into a database for hierarchical cluster analysis to construct clustering schemes. For each pair of factors, a “distance” score was computed. Four types of hierarchical cluster analysis were performed. Focus groups were then used to validate the clustering scheme. The resulting taxonomy was independently reviewed by two qualified individuals in ethics.

Design: A mixed methods study using nonexperimental research with a convergent-parallel approach and a data-transformation design

Figure 15.3 Data-Validation Design



Note: Any research design designated as experimental, quasi-experimental, or nonexperimental research can be used for the QUAN phase, and cross sections of qualitative data can be collected for the qual phase.

Example for Figure 15.3

Abrams, L. S., Shannon, S. K. S., & Sangalang, C. (2008). Transition services for incarcerated youth: A mixed methods evaluation study. *Children and Youth Services Review, 30*, 522–535.

Research Aims: Examine recidivism outcomes for youth participants in a transitional living program at 1-year postrelease; explore child protective services involvement as a risk factor for recidivism at 1-year postrelease; compare youth and staff perspectives on the strengths and limitations of the transitional living program in preparing youth for community reentry.

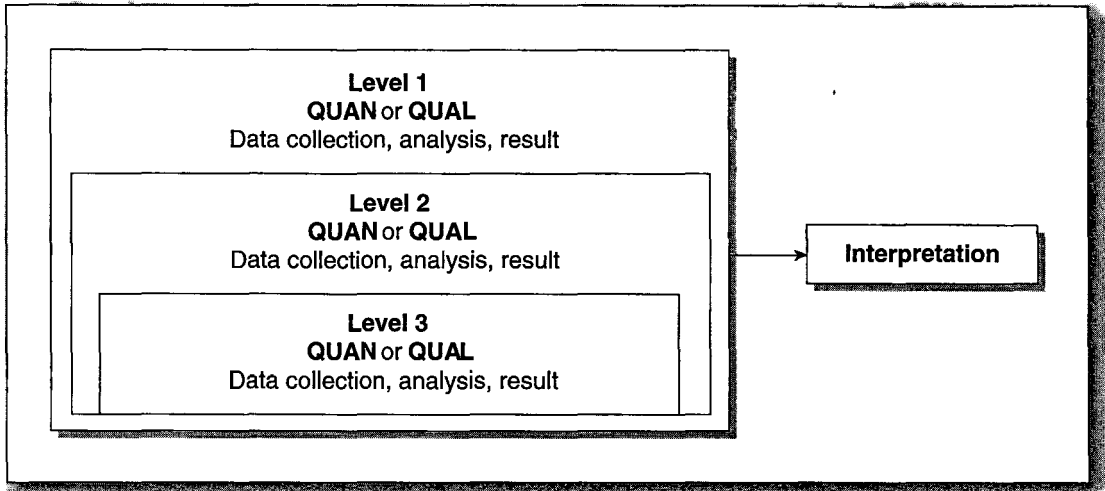
Research Question (Mixed Method): To what extent do the open-ended themes support the survey results?

Procedures

Quantitative: Archival data were retrieved from the state administrative data system for juvenile and adult offenders and official client case records. Variables retrieved from the state administrative data included basic demographic information, number of prior arrests, new substantiated crimes up to 1-year postrelease, and participation in the transition living program. Variables retrieved from the intake forms included a history of child welfare system involvement, family structure, substance abuse, and additional comprehensive descriptors of each case. The primary independent variable was participation in the 6-week transition living program. Child welfare system involvement was another independent variable. The dependent variable for the analysis was recidivism at 1-year postrelease. Control variables retrieved from the administrative database included number of prior arrests, race, and age at admission to program.

Qualitative: The sample for the qualitative component of the study included 10 youth who participated in the transition living program and were interviewed repeatedly over a 6-month period. Interviews with youth were semistructured. The interviews were taped with a digital recording device. Interviews with staff occurred after the youth-interview component of the project was completed. Questions were geared to gather staff perspectives on the important components of transition, the benefits and limitations of the transition program, and their views on the challenges of youths' postrelease environments. Digitally taped interviews were transcribed verbatim and further reduced via data analysis.

Design: A mixed methods study using nonexperimental research with a convergent-parallel approach and a data-validation design

Figure 15.4 Multilevel Design

Note: Any research design designated as experimental, quasi-experimental, or nonexperimental can be used for the QUAN phases, and any approach designated under the qualitative method can be used for the QUAL phases.

Example for Figure 15.4

Elliott, M. S., & Williams, D. I. (2002). A qualitative evaluation of an employee counselling service from the perspective of client, counselor, and organization. *Counselling Psychology Quarterly*, 15(2), 201–208.

Research Questions

Quantitative and Qualitative: What is the degree of congruence across the aims and expectations of all three parties (the client, counselor, and organization)? Are their needs being met?

Mixed Method: What similarities and differences exist across levels of analyses?

Procedures: Data were collected from three different levels: (a) the client, (b) the counselor, and (c) the organization. Client data included semistructured interviews with each party. The form of the interviews followed the client's path from before counseling to the present. Participants were interviewed between 6 and 20 months posttermination of counseling. The counselor completed an "encounter questionnaire" with respect to each of the clients interviewed in the study. The counselor was asked for an evaluative

trace through counseling for each client, concentrating on relevant areas such as conceptualization, process and outcome, impact on the organization, and professional and personal issues raised by the work. An intense in-depth study of the counselor perspective was conducted through a series of four linked interviews with one of the counselors. Organizational data were also collected from semistructured interviews with high-level officials within the organization. Awareness was assessed by administering questionnaires to a stratified sample of the staff. Bottom-line benefits were assessed by reviewing the sickness records of the client group.

Design: A mixed methods study using nonexperimental research with a convergent parallel approach and a multilevel design

Reviewing the Content and Testing Your Knowledge

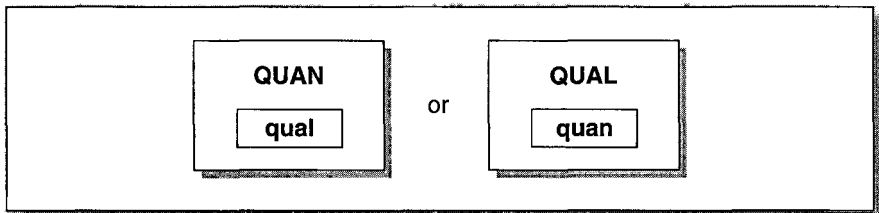
Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Mixed Method** and the **Convergent-Parallel Approach**. The research will be considered nonexperimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.
3. Develop the appropriate primary research question to be associated with this design (including both QUAL and QUAN questions).
4. Discuss the sampling strategy and technique to be used.
5. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures.
6. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
7. Briefly discuss the strengths and limitations associated with this approach and the specific design.

CHAPTER 16

EMBEDDED APPROACH



The embedded approach is a *nested* approach and is used when one type of data (QUAN or QUAL) is most critical to the researcher. This approach is used when different questions require different types of data (qualitative and quantitative). The embedded approach is appropriate when one type of data clearly plays a secondary role and would not be meaningful if not embedded within the primary data set. The embedded approach is also useful when the researcher logistically cannot place equal priority on both types of data or simply has little experience with one of the forms of data. Many variants of the embedded approach have been proposed, such as the embedded narrative and ethnographic designs. For many years, clinical psychologists have used a form of the embedded narrative case study design for cases classified in abnormal psychology (see Oltmanns et al., 2014)—that is, a clinician would collect relevant quantitative indices and qualitative data and develop a cohesive narrative account, explaining the clinical features of the individual case. Based on theoretical

and logistical considerations, many other design variants of the embedded approach can be meshed with many of the traditional designs presented in this book. These are sometimes referred to as hybrid designs.

♦ EMBEDDED-EXPERIMENT DESIGN

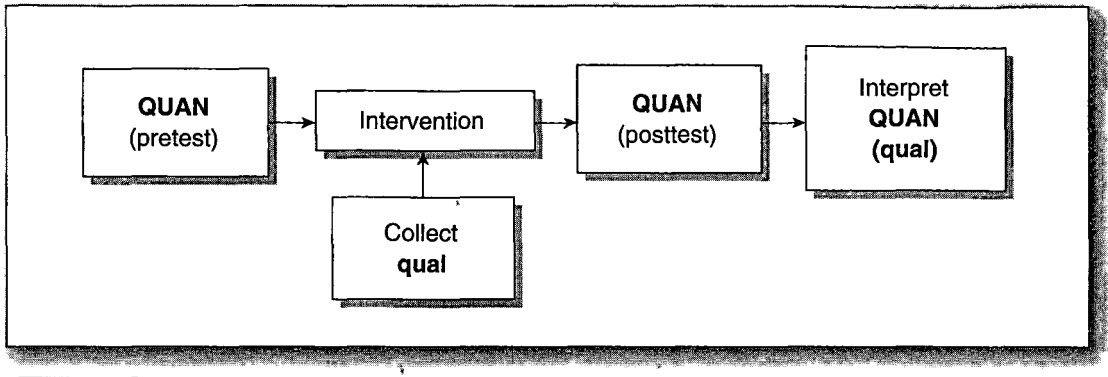
The embedded-experiment design allows the researcher to embed qual data within experimental research. If the research is considered quasi-experimental, then the design can be referred to as an embedded quasi-experimental design. The researcher can use any research design that is designated as such (i.e., within- or between-subject approaches). This model can be further designated as one-phase, which allows the researcher to collect the qual data during the intervention, or two-phase, in which the researcher collects qual data before and after the experimental or quasi-experimental phase.

♦ EMBEDDED-CORRELATIONAL DESIGN

The embedded-correlational design allows the researcher to embed qual data within nonexperimental research (observational approach). The designs can be either predictive or explanatory. QUAN data is the emphasis within this design.

♦ EMBEDDED CASE STUDY DESIGN

The embedded case study design can be applied as a means to explore a phenomenon within its real-world context when the boundaries between phenomena and context are not clearly evident in which multiple data sources and types (QUAL and quan) are used. QUAL data is the emphasis, while the quan data provides a supplementary role to the qualitative findings. An ethnographic or narrative approach is commonly applied to guide the tenets of the case study design.

Figure 16.1 Embedded-Experiment Design (One-Phase)

Note: Any research design designated as a between-subjects or repeated-measures approach can be used for the QUAN phase, and a cross section of qualitative data can be collected for the qual phase.

Example for Figure 16.1¹

Zydney, J. M. (2008). Cognitive tools for scaffolding students defining an ill-structured problem. *Journal of Educational Computing Research*, 38(4), 353–385.

Research Questions

Quantitative and Qualitative: What is the effect of cognitive tools in scaffolding students defining an ill-structured problem, as measured by (a) students' problem understanding, (b) ability to generate questions, and (c) ability to formulate hypotheses on how to solve the problem?

Mixed Method: How do the qual results inform the development of the treatment? What additional information is obtained during the trial from the qual data? How do the qual results expand on the QUAN data?

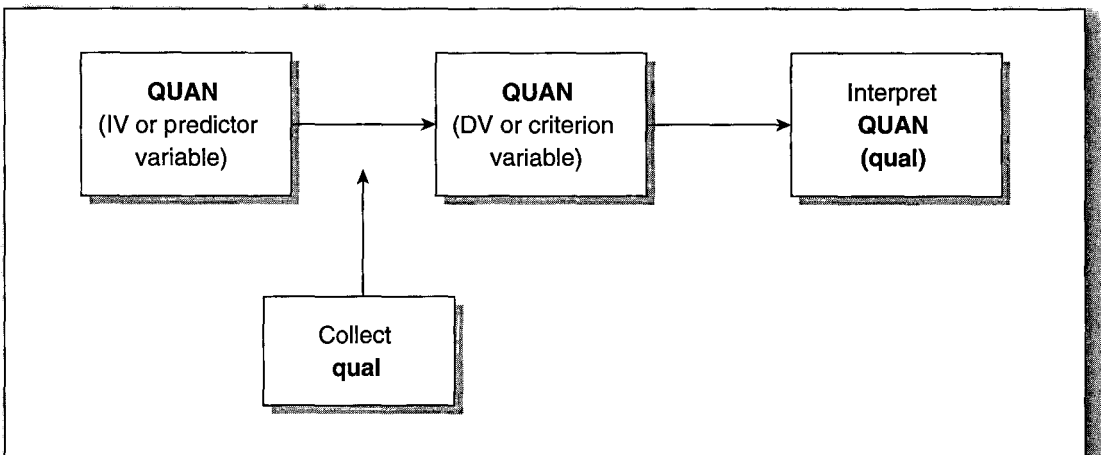
Procedures: During the first session, students were introduced to the project and completed the fluency test. During the second session, students met

¹The example uses the one-phase approach to the embedded design (also known as a *concurrent nested mixed methods design*) with the qualitative data collected during the intervention. A two-phase approach can also be used when the qualitative data is collected before or after the intervention.

their supervisor and their client, who explained the acid rain problem faced by the company. During the third session, students discussed what information they needed and saw a brief demonstration on how to find information within the *Pollution Solution* learning environment. During the fourth session, students independently researched and took notes about the problem. The students with the higher-order thinking tool and the combination tool completed their status reports, and students in the organization tool and control groups continued their research. After completing their research plans, the students completed a questioning assessment. After this assessment, the students answered a few survey questions to determine if they were absent at any time, worked at home, lost any data during the study, or discussed their work outside of class.

The quantitative and qualitative data were collected concurrently (when data are collected during the intervention, it is considered a one-phase approach). The quantitative data were obtained through computer-based assessments, and the qualitative data were captured through classroom observations. The classes were randomly assigned to one of four conditions of the treatment. The control group provided students with directions to write the research plan. Another group received the organization tool. Field notes carefully described the classroom activity during the study, and special notes were made of any differences between the classes. To analyze these field notes, qualitative differences among the classes were coded. These qualitative differences were organized in a chart along with

Figure 16.2 Embedded-Correlational Design



Note: The observational design can be either explanatory or predictive.

the quantitative findings in order to see whether the statistical differences could be explained by the qualitative differences among the classes.

Design: A mixed methods study using experimental research with an embedded approach and an embedded-experiment design

Example for Figure 16.2

Forchuk, C., Norman, R., Malla, A., Martin, M., McLean, T., & Cheng, S., . . . Gibney, C. (2002). Schizophrenia and the motivation for smoking. *Perspectives in Psychiatric Care*, 38(2), 41–49.

Research Questions

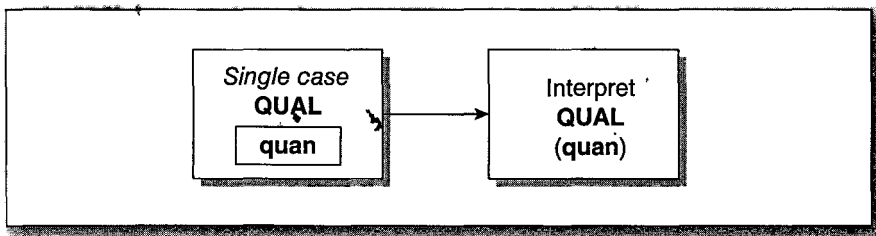
Quantitative and Qualitative: Are individuals with schizophrenia motivated to smoke for relief of psychiatric symptoms and to relieve anti-psychotic medication side effects?

Mixed Method: How does the qual data add to the expansion of the constructs in the correlational model?

Procedures: Participants were selected randomly by staff members at various settings. After informed consent was obtained, the interviewer and participant agreed on a meeting place. The interview took approximately one hour. All data were collected through an interview format, which included the questionnaires and open-ended questions.

Data were collected through a single interview with individuals who have been diagnosed with schizophrenia for a minimum of 1 year and who smoke. The research design included two components to examine smoking

Figure 16.3 Embedded Case Study Design



Note: Yin (2009) refers to the term *embedded* as collecting data from multiple sources regardless of method. "Embedded" in this example refers to the collection of quan data within the QUAL data collection procedures.

and schizophrenia: (a) a descriptive, correlational design that described and examined the relationships among psychiatric symptoms, medication side effects, and reasons for smoking; and (b) a content analysis of open-ended questions related to the subjective experience of smoking.

Design: A mixed methods study using nonexperimental research with an embedded approach and an embedded-correlational design

Example for Figure 16.3

Beeke, S., Wilkinson, R., & Maxin, J. (2007). Individual variation in agrammatism: A single case study of the influence of interaction. *Journal of Language and Communication Disorders*, 42(6), 629–647.

Research Questions

Quantitative and Qualitative: In what ways do grammatical structure, verbs, and argument structure qualitatively reveal the single case? What is the influence of the interaction on the structure of the utterances from the case?

Mixed Method: How does the quan data add to and expand on the QUAL findings?

Procedures: A single case diagnosed with severe and chronic agrammatism is videotaped completing a set of spoken tests. Videos were also taken of independent conversations between the case and family members. The case was administered six different measures, which were qualitatively driven but contained quantitative components. These include the Psycholinguistic Assessment of Language Processing in Aphasia, Subtest 53: Spoken Picture Naming (PALPA), Thematic Roles in Production (TRIP), Verb and Sentence Test (VAST), Cookie Theft Picture Description, Dinner Party Cartoon Strip Description, and the Cinderella Story-Telling technique. The PALPA, TRIP, and VAST were analyzed and assigned quantitative values based on the outcome data. The data from the remaining assessments were microanalyzed based on all the utterances in each set. The data were then combined and discussed as a whole using a narrative approach.

Design: A mixed methods study using nonexperimental research with an embedded approach and an embedded case study design

Reviewing the Content and Testing Your Knowledge

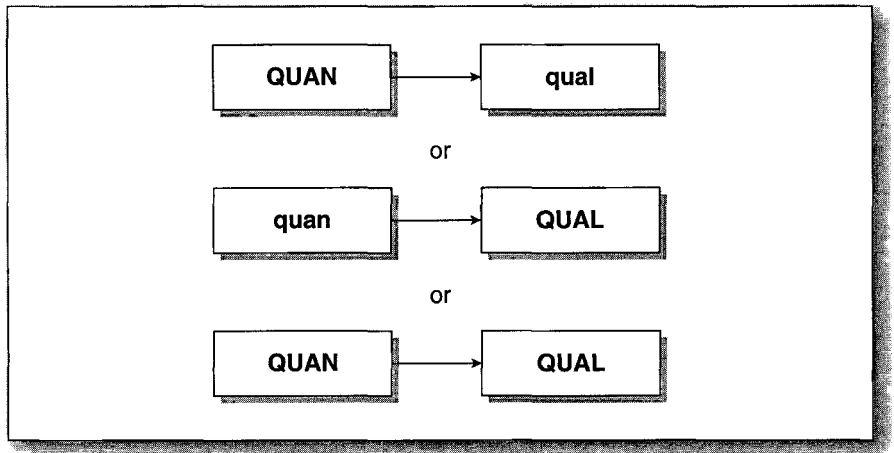
Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Mixed Method** and the **Embedded Approach**. The research will be considered nonexperimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.
3. Develop the appropriate primary research question to be associated with this design.
4. Discuss the sampling strategy and technique to be used.
5. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures.
6. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
7. Briefly discuss the strengths and limitations associated with this approach and the specific design.

CHAPTER 17

EXPLANATORY- SEQUENTIAL APPROACH



The explanatory-sequential approach is a *sequential* approach and is used when the researcher is interested in following up the quantitative results with qualitative data. Thus, the qualitative data is used in the subsequent interpretation and clarification of the results from the quantitative data analysis. In many instances, because the QUAN design is the emphasis, a generic qual design is used in explanatory approaches. This two-phase approach is particularly useful for a researcher interested in

explaining the findings from the first phase of the study with the qualitative data collected during Phase 2. However, either the qualitative or quantitative data (or both equally) may be the primary focus of the study (see introductory figure). For example, the qualitative phase is often emphasized when using the participant-selection design.

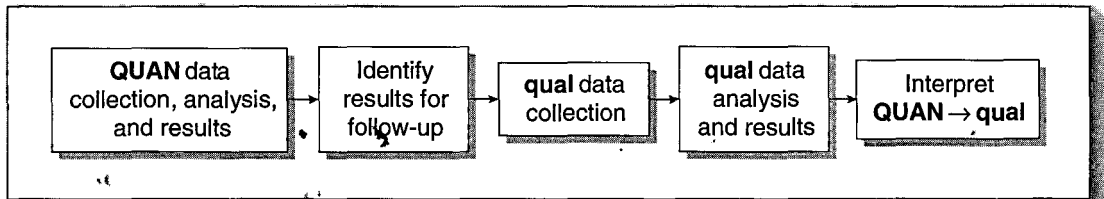
FOLLOW-UP EXPLANATIONS DESIGN ♦

The follow-up explanations design provides a framework for the researcher to collect qual data in order to expand on the QUAN data and results. Within this design, a researcher analyzes the relevant QUAN results and then uses the qual findings to further explain the initial QUAN results. Thus, the primary emphasis is on the QUAN results.

PARTICIPANT-SELECTION DESIGN ♦

The participant-selection design involves a two-phase process: First, the participant selection (Phase 1) is conducted using a quantitative method, followed by a qualitative data collection phase (Phase 2). Participants are selected during the first phase based on parameters set a priori by the researcher as a means of purposeful sampling. Thus, the quan phase is strictly used to generate the sample.

Figure 17.1 Follow-Up Explanation Design



Note: Any research design designated as experimental, quasi-experimental, or nonexperimental research can be used for the initial QUAN phase of this design, and cross sections of qualitative data can be collected for the qual phase.

Example for Figure 17.1

Lee, K. S., Osborne, R. E., Hayes, K. A., & Simoes, R. A. (2008). The effects of pacing on the academic testing performance of college students with ADHD: A mixed methods study. *Journal of Educational Computing Research*, 39(2), 123–141.

Research Questions

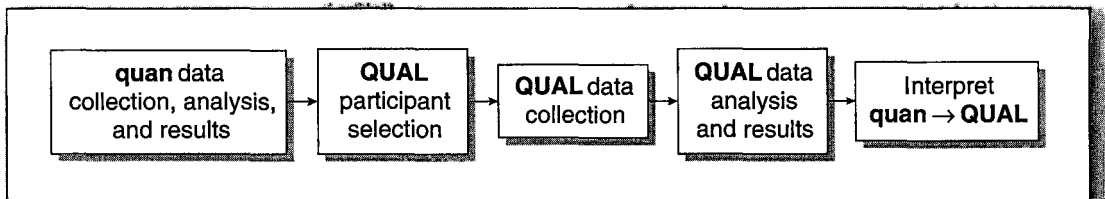
Quantitative and Qualitative: What is the relationship between computer-paced and student-paced item presentation on the academic test performance in college students diagnosed with attention-deficit/hyperactivity disorder (ADHD)?

Mixed Method: In what ways do the qual data help to explain the QUAN results?

Procedures: Participants were randomly assigned to one of two treatment conditions. In the computer-paced testing condition, the students were allowed 90 seconds per question and were forced to move on to the next question when the time expired. In the student-paced testing condition, students were allowed an average of 90 seconds per question but were not forced to move on to the next question. Upon completion of either the computer-paced or student-paced test, each participant was individually interviewed face to face by the primary investigator to explore the student's perception of the testing experience.

This exploratory study used a follow-up explanation with quasi-experimental design (QUAN) to explore and explain the effects of paced-item presentation for college students diagnosed with ADHD. The goal was to analyze two testing conditions and interpret their impact on a small number of participants who participated in the study.

Figure 17.2 Participant-Selection Design



Note: Cross sections of quantitative data can be collected for the initial quan phase, and any approach and design designated under the qualitative method can be used for the QUAL phase.

Design: A mixed methods study using quasi-experimental research with an explanatory-sequential approach and a follow-up explanations design

Example for Figure 17.2.

Hannan, M., Happ, M. B., & Charron-Prochownik, D. (2009). Mothers' perspectives about reproductive health discussion with adolescent daughters with diabetes. *The Diabetes Educator*, 35(2), 265–273.

Research Aim and Question

Quantitative and Qualitative: Explore mothers' perspectives about reproductive health (RH) discussions with their adolescent daughters with diabetes.

Mixed Method: Which cases provide the best insights into the quan results?

Procedures: The researchers followed a two-phase sequential explanatory process. Phase 1 involved the selection of participants from a larger study sample through purposeful sampling. More specifically, criterion-related purposeful sampling was used to select a subset of 10 mothers from the total sample. Participants were selected for variation on the following parameters: baseline knowledge (modified Family Planning Behavior and Diabetes Study questionnaire) and intention scores (modified Initiating Discussion questionnaire), daughters' group assignment (IG or CG), and daughters' age group. Data from questionnaires administered in Phase 1 were used for criterion-related sampling.

Phase 2, the focus of the report, was a qualitative descriptive study using open-ended, semistructured telephone interviews. The principal investigator conducted interviews via the telephone using a semistructured interview guide that began with a grand tour question ("I'd like for you to tell me about discussing RH issues, such as monthly periods, sex, birth control, or pregnancy, with your daughter"). Additional questions probed the mother's perceptions of who initiated discussions, timing of discussions, barriers and facilitators to discussions, and her comfort with initiation of discussion. There was no time limit to the interviews, which generally lasted 20 to 30 minutes. Interviews were digitally recorded, transferred to a secure laptop, transcribed verbatim, and reviewed for accuracy. The interviews were conducted over a 1-year period of time, which began 2 years after mothers' completion of the quantitative portion of this study. Qualitative content analysis techniques were used to analyze the interview transcripts.

Design: A mixed methods study using nonexperimental research with an explanatory-sequential approach and a participant-selection design

Reviewing the Content and Testing Your Knowledge

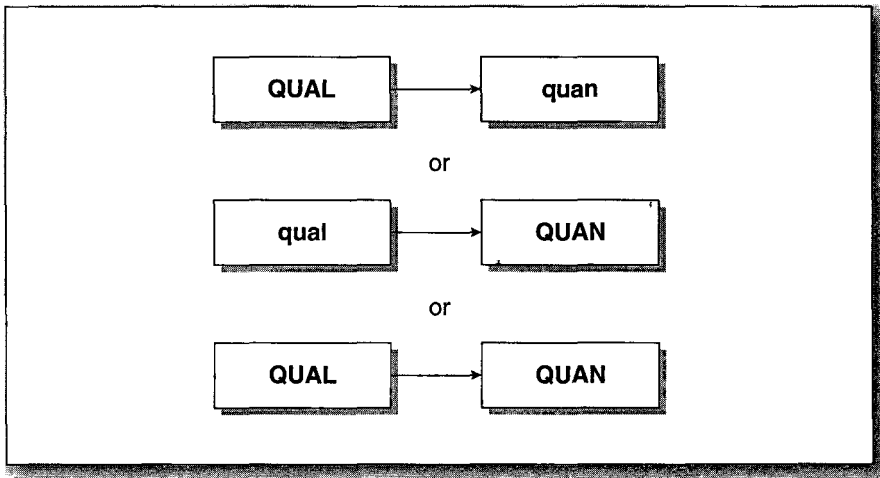
Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Mixed Method** and the **Explanatory-Sequential Approach**. The research will be considered nonexperimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.
3. Develop the appropriate primary research question to be associated with this design.
4. Discuss the sampling strategy and technique to be used.
5. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures.
6. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
7. Briefly discuss the strengths and limitations associated with this approach and the specific design.

CHAPTER 18

EXPLORATORY- SEQUENTIAL APPROACH



The exploratory-sequential approach is a *sequential* approach and is used when the researcher is interested in following up qualitative findings with quantitative analysis. This two-phase approach is particularly useful for a researcher interested in developing a new instrument, taxonomy, or treatment protocol (Creswell & Plano Clark, 2011). The researcher uses the qualitative (exploratory) findings from the first phase to help develop the instrument or treatment and then tests this product during the second phase (quantitative). In general, when variables are unknown,

this approach is useful to identify important variables (Phase 1) for subsequent quantitative analysis (Phase 2). It is also a useful approach for revising existing instruments and treatment protocols, as well as for developing and testing a theory. Although the QUAL phase is usually the primary focus, either the qualitative or quantitative phase (or both equally) may be the primary emphasis of the study (see introductory figure).

♦ INSTRUMENT-DEVELOPMENT DESIGN

The instrument-development design is often QUAN emphasized and provides a framework for the researcher to first develop and then test (psychometrically) an instrument on a specific population. With this design, the researcher uses the qualitative results to help construct the instrument and validates the instrument during the subsequent quantitative phase. Either the qualitative or quantitative data (or both equally) may be the primary emphasis of the study.

♦ THEORY-DEVELOPMENT DESIGN

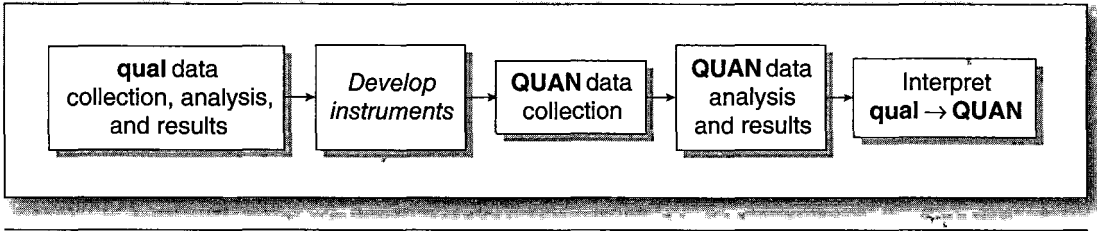
The theory-development (and taxonomy-development) design is often QUAL emphasized. The researcher uses the qualitative data collected during the first phase to identify, develop, and construct a classification system or theory. The taxonomy or theory is subsequently analyzed quantitatively during Phase 2. Oftentimes, researchers will use the qualitative findings to develop their research questions, which guide the quantitative phase of the study. Either the qualitative or quantitative data (or both equally) may be the primary emphasis of the study.

♦ TREATMENT-DEVELOPMENT DESIGN

The treatment-development design is both QUAL and QUAN emphasized and provides a framework for the researcher to develop and then test a treatment protocol or approach with a specific population. With this design, the researcher uses the qualitative results to help construct the treatment

protocol and then tests the efficacy of the treatment during the subsequent quantitative phase. Either the qualitative or quantitative data (or both equally) may be the primary emphasis of the study.

Figure 18.1 Instrument-Development Design



Note: Cross sections of qualitative data can be collected for the qual phase, and an explanatory design within the observational approach is typically used for the QUAN phase of this design.

Example for Figure 18.1

Zolotor, A. J., Runyan, D. K., Dunne, M. P., Jain, D., Petrus, H. R., Ramirez, C., . . . Isaeva, O. (2009). ISPCAN Child Abuse Screening Tool Children's Version (ICAST-C): Instrument development and multi-national pilot testing. *Child Abuse & Neglect*, 33(11), 833–841.

Research Aims and Question

Phase 1: Develop a child victimization survey.

Phase 2: Examine the performance of the instrument through a set of international pilot studies.

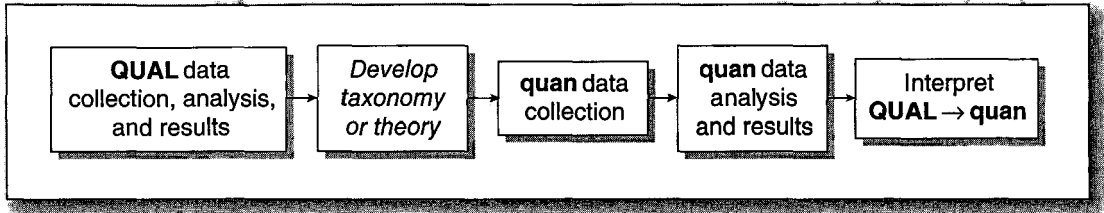
Mixed Method: What items and scales represent the qual findings?

Procedures: The researchers developed the initial draft of the instrument after receiving input from scientists and practitioners representing 40 countries. The original instrument contained 82 screener questions regarding the potentially victimizing experiences at home and school or work. Volunteers from the larger group of scientists participating in the Delphi review of the ISPCAN Child Abuse Screen Tool–Parent Version (ICAST-P) and Retrospective Version (ICAST-R) reviewed the Children's Version (ICAST-C) by e-mail in two rounds, resulting in a final instrument. The ICAST-C was then translated and back-translated into six languages and field tested in four countries

using a convenience sample of 571 children 12 to 17 years of age, who were selected from schools and classrooms to which the investigators had easy access.

Design: A mixed methods study using nonexperimental research with an exploratory-sequential approach and an instrument-development design

Figure 18.2 Theory-Development Design



Note: Any approach and design designated under the qualitative method can be used for the QUAL phase, and cross sections of quantitative data can be collected for the quan phase.

Example for Figure 18.2

Kartalova-O'Doherty, Y., & Doherty, D. T. (2008). Coping strategies and styles of family carers of persons with enduring mental illness: A mixed methods analysis. *Scandinavian Journal of Caring Sciences*, 22(1), 19–28.

Research Aims and Question

Phase 1: Identify and describe the coping strategies and styles of participants.

Phase 2: Investigate the interaction of identified coping strategies and styles with the sociodemographic characteristics of participants, and then explore the interaction of the identified coping styles with contextual factors.

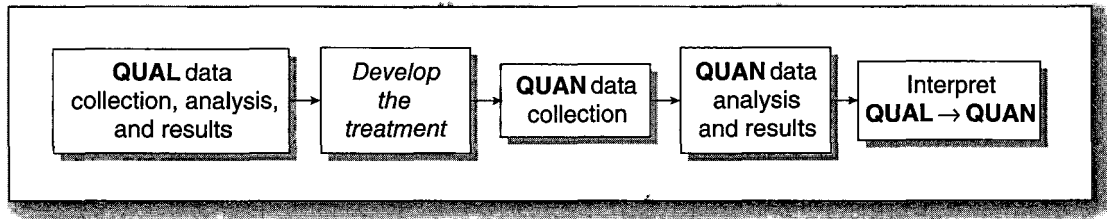
Mixed Method: How do the quan results generalize to the QUAL findings?

Procedures: During Phase 1, content analysis was used to explore and classify the self-reported coping strategies and styles of participants as emerging from the original interviews. Statistical procedures were employed during Phase 2 to explore the interaction of coping styles with contextual

factors, including duration of illness, living arrangements, and occupational status of the ill relative. The mixed methods analysis allowed the researchers to explore the qualitative data from the perspective of well-established quantitative findings in the areas of stress and coping and to further interpret the emergent findings using qualitative data.

Design: A mixed methods study using nonexperimental research with an exploratory-sequential approach and a theory-development design

Figure 18.3 Treatment-Development Design



Note: Any approach and design designated under the qualitative method can be used for the QUAL phase, and any research design designated as experimental, quasi-experimental, or nonexperimental research can be used for the initial QUAN phase of this design.

Example for Figure 18.3

Nagel, T., Robinson, G., Condon, J., & Trauer, T. (2009). Approach to treatment of mental illness and substance dependence in remote indigenous communities: Results of a mixed methods study. *Australian Journal of Rural Health, 17*(4); 174–182.

Research Aims and Question

Phase 1: Develop a culturally adapted brief intervention for indigenous people with chronic mental illness.

Phase 2: Evaluate the efficacy of the brief intervention.

Mixed Method: What treatment was developed from the QUAL findings?

Procedures: An exploratory phase of qualitative research was followed by a nested randomized controlled trial. The first phase of the study focused on understanding local perspectives of mental health through collaboration

with local aboriginal mental health workers (AMHWs). These perspectives were then incorporated into a brief motivational care planning (MCP) intervention. Qualitative data providing rich description of the personal experiences of patients were gathered concurrently with the randomized controlled trial and integrated into the final analysis.

The exploratory phase (Phase 1) of the study was conducted over 12 months. Data were collected during 15 field trips of 1- to 3-day duration. Group and individual interviews were supplemented by informal observation. Three key themes emerged: the importance of family, the strength gained from traditional and cultural activities, and the importance of a storytelling approach to sharing information.

During Phase 2, 49 indigenous patients with mental illness and 37 caregivers were recruited. Patient participants were randomly allocated to two groups using a block-randomization, random-number-sequence technique after completion of baseline measures. Participants, caregivers, and AMHWs were given an explanation of the project in spoken, written, and pictorial format. When necessary, translation to local language was provided by the AMHWs in order to ensure that informed consent was obtained. The treatment was delivered at baseline in the first group (the “early treatment” group) and at 6 months in the second group (the “late treatment” group).

Design: A mixed methods study using both nonexperimental and experimental research with an exploratory-sequential approach and a treatment-development design

Reviewing the Content and Testing Your Knowledge

Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Mixed Method** and the **Exploratory-Sequential Approach**. The research will be considered nonexperimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.

3. Develop the appropriate primary research question to be associated with this design.
4. Discuss the sampling strategy and technique to be used.
5. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures.
6. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
7. Briefly discuss the strengths and limitations associated with this approach and the specific design.

CHAPTER 19

MIXED METHODS, CASE STUDIES, AND SINGLE-CASE APPROACHES

The primary reason for using mixed methods is to maximize the use of blending methods to answer research questions within a study (i.e., converge and confirm results from different methodological techniques). Keep in mind that the use and application of mixed methods in education and the social and behavioral sciences are still relatively new and evolving (see Tashakkori & Teddlie, 2010b). Many of the designs presented are difficult to locate in the literature; that is, authors typically do not indicate the name of the mixed method research design (or use different names) in their published manuscripts. Nonetheless, there is a growing interest and need for the application of mixed methods as a means to reveal complex and relevant scientific inquiries. There are many applications of mixed methods not yet identified in the literature or in textbooks that we propose. Based on our observations in the field, we recommend combining qualitative methodology with the family of A-B designs (i.e., the single-case approach). Developing a structure and framework for mixed method single-case approaches can strengthen the results from $N = 1$ designs en route to implying causal relations.

The mixed method single-case approach still maintains the key characteristics (as defined by the quantitative methodological tenets), which are (a) continuous assessment (repeated measures), (b) baseline assessment,

(c) accounting for stability in performance, and (d) the introduction of varied phases. However, the qualitative (qual) method should serve as a secondary role to the quantitative (QUAN) method (i.e., the emphasis is on the design of the single-case approach). Because the qualitative method is secondary, a sound generic qualitative design should usually suffice for these applications. For example, a cross section of qual data can be collected concurrently, sequentially, or can be nested (embedded) within the design. When applying these designs and staying true to the tenets of mixed methodology, it is critical to discuss how the qual findings add to, explain, and expand on the QUAN results.

MIXED METHOD A-B-A DESIGNS ♦

Although we present diagrams of the A-B-A mixed method design, any version of the A-B family of designs can be used (e.g., A-B-A-B, A-B-C, multiple baseline, changing criterion, etc.). Diagram 19.1 is the A-B-A concurrent design. The qual data is collected concurrently (simultaneously) throughout the process of the design. Next, Diagram 19.2 illustrates the A-B-A sequential design. Within this design, qual data is collected sequentially throughout the various phases of the treatment and baseline applications. In addition to collecting qual data between each session, the qual data can also be collected prior to and after the baseline and follow-up sessions. Last, Diagram 19.3 is an A-B-A nested design. This design allows the researcher to collect qual data prior to and following the application of the entire design. See the diagrams for diagrammatic representations of the mixed methods single-case approaches. The applications of these designs should always be based on theoretical and logistical considerations.

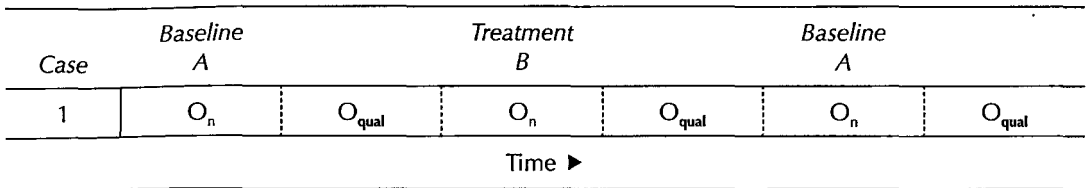
Diagram 19.1 A-B-A Concurrent Design

Case	Method	Baseline A	Treatment B	Baseline A
1	QUAN	O _n	O _n	O _n
	qual	O _{qual}	O _{qual}	O _{qual}

Time ►

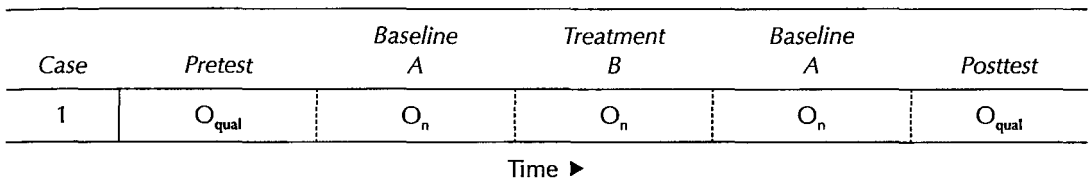
Note: Multiple forms and types of qual data can be collected during each phase.

Diagram 19.2 A-B-A Sequential Design



Note: A cross section of qual data can be collected at any point between treatment and baseline phases.

Diagram 19.3 A-B-A Nested Design

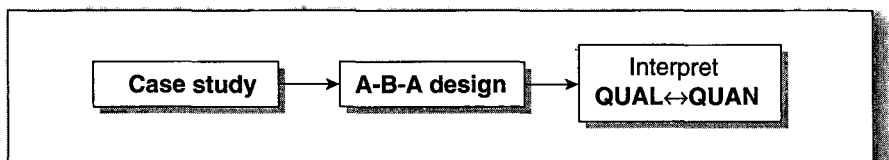


Note: A cross section of qual data is collected prior to and following the completion of the study.

Sequential Case Study Single-Case Design

As seen in Figure 19.1, we propose the application of combining the case study design and the single-case approach as a means to provide a truly in-depth and rigorous analysis and assessment of a single participant ($N = 1$). This design would be considered an exploratory approach, and the emphasis would be both on qualitative (QUAL) and quantitative (QUAN) methods sequentially delivered. This design is applicable in a wide array of disciplines, such as education, psychiatry, rehabilitation, and medicine. The general steps would include, for example, using the case study approach

Figure 19.1 Sequential Case Study Single-Case Design



Note: Any version of the case study (see Appendix B) can be used as defined by Yin (2013) or Creswell (2014), and any version of the A-B family of designs can be used.

to detail and reveal the intricate cognitive and behavioral patterns associated with a child diagnosed with a pervasive developmental disorder. For example, information gathered from the case study can be applied to a form of cognitive behavioral therapy and the effects assessed through the use of an A-B-A design. As with all mixed methods studies, the results from the case study and A-B-A design should be analyzed and discussed individually and collectively.

Reviewing the Content and Testing Your Knowledge

Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Mixed Method A-B-A Design**. The research will be considered experimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.
3. Determine if the mixed method aspect is concurrent, sequential, or nested.
4. Develop the appropriate primary research question to be associated with this design.
5. Discuss the sampling strategy and technique to be used.
6. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures. Be sure to include both the qualitative and quantitative data collection procedures.
7. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
8. Briefly discuss the strengths and limitations associated with this approach and the specific design.

CHAPTER 20

ACTION RESEARCH APPROACHES

The action research approach is a form of research that enables individuals to reveal functional solutions to problems encountered in the context in which they operate or work. The general framework (which can be likened to the scientific method) includes a cyclical approach, starting with the identification of the problem, data collection, analysis, and then a feedback phase. Chevalier and Buckles (2013) discussed action research as a spiral of activity, including the phases of *plan*, *act*, *observe*, and *reflect*. Stringer (2013) presented the action research model in three phases: *look*, *think*, and *act*. Action research is best suited for educational, health, and community organizations in which the intent is to generalize the findings back to the sample and context where the research takes place, as opposed to extending the findings to the population. We would like to point out that many texts on action research claim that action research is not like traditional scientific inquiry because the findings are not intended to be generalized (i.e., external validity) to populations outside of the research focus. However, the traditional application of the scientific method and experimental research is primarily intended to ensure aspects of internal validity (i.e., the findings can be attributed to the program or treatment).

We have also found in reviewing many action research texts that authors provide the general cyclic framework for action research then discuss a mixed bag of quantitative and qualitative data collection strategies

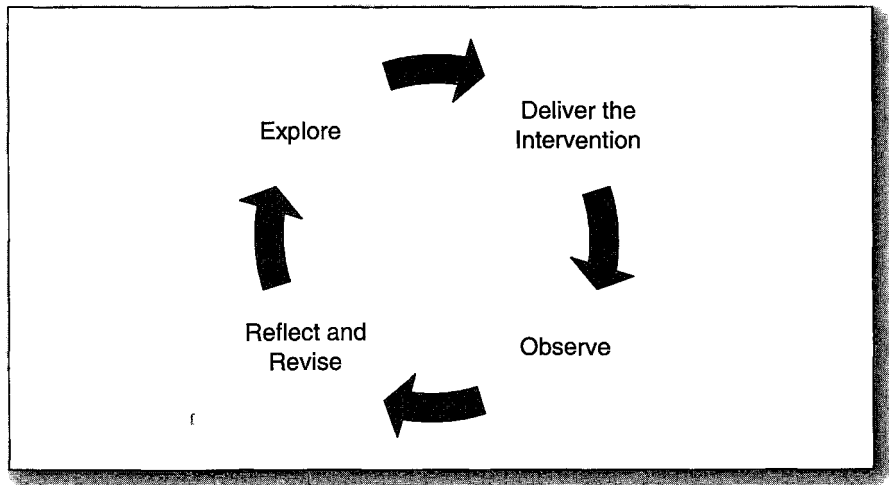
as a means to answer (or solve) the stated problems. Using mixed methods for action research is advisable, but it is vital to include the data collection strategies as part of a greater research-design framework to ensure the results can ultimately be attributed to the program or treatment of interest. Therefore, when applying the action research approach it is advisable to follow the steps of the scientific method and use the sound principles and application of research designs that are known to produce results while ensuring adequate levels of internal validity.

Where action research begins to diverge from the traditional form of experimental research is with the role of the researcher. The researcher is considered more or less a facilitator or consultant to the stakeholders who generally drive the research-question process, although this is clearly not true in all circumstances. Nonetheless, this, in many ways, is more like the program evaluation model, only on a smaller scale. This lends itself to what is known as the participatory action research (PAR) model, in which the research is based on community-driven goals, and all those affected participate and take action. This is often referred to as *community-based action research*. The action research approach is an attempt to involve the interest of those affected and who are concerned about problems by providing a framework that delivers workable solutions.

We should note that the action research approach is still evolving, and new ideas and frameworks are often introduced and refined to suit specific research contexts. For example, researchers have proposed and demonstrated that the application of collaborative and analytic autoethnography can be combined with the action research approach (Acosta, Goltz, & Goodson, 2015). PAR models have also been combined with the Delphi technique as a means to address specific research objectives (see Fletcher & Marchildon, 2014). Helmer (1967) originally developed the Delphi technique as a systematic approach to garner relevant and appropriate data from those who are considered experts in a respective field in the absence of a standard theoretical framework. This is a complementary approach for the PAR model and should be used more often in relevant contexts. We refer the reader to learn more about the application of the Delphi technique in the health sciences (Keeney, McKenna, & Hasson, 2011) and in education (Manley & Zinser, 2012).

The action research approach is iterative and cyclical and can be represented as the following characteristic cycle:

- Explore
- Deliver the intervention
- Observe
- Reflect and revise

Figure 20.1 The Cyclical Action Research Approach

We refer the reader to the following books to learn more about the action research approach:

Chevalier, J. M., & Buckles, D. J. (2013). *Participatory action research: Theory and methods for engaged inquiry*. London, England: Routledge.

Stringer, E. T. (2013). *Action research* (4th ed.). Thousand Oaks, CA: Sage.

♦ EMBEDDING RESEARCH DESIGNS WITHIN THE ACTION RESEARCH APPROACH

We recommend four basic research designs that can be embedded within the action research approach. Because this approach is indeed cyclical and participatory in nature, it does not preclude it from including relevant aspects of the research design framework to secure the internal validity of the findings. That is, the researcher can confidently attribute the findings to the program or treatment under examination and reduce the probability of reporting spurious results. Keep in mind that the findings are intended to feed back into the environment from which the research takes place, and they are not intended to generalize to the greater population of interest.

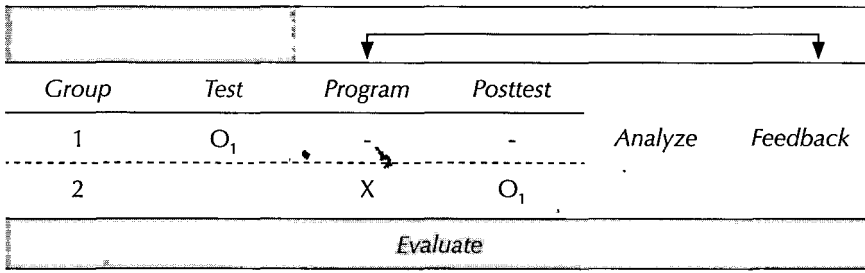
We detail the components of a design used within the action research approach in Table 20.1. Dependent variables (O_1) should be indicated after

the program of interest (X) is targeted. Various forms of qualitative and quantitative data should be collected. The types of data depend on data that are available and relevant to the stated research questions or objectives. Data can be collected, for example, via interviews, focus groups, observations, assessments, documents, surveys, and reports. Throughout the process of analyzing and feeding the data back into the program, an evaluation process should be ongoing. The evaluation process ensures fidelity of the procedural steps. Ideally, an individual who is external to the process should be charged with the task of evaluation to minimize the bias and overall workload.

Table 20.1 Design Notations for the Action Research Approach

<i>Design Notation</i>	<i>Design Element</i>
O_n	The dependent variable (includes qual and quan data collected from artifact, observational and inquiry data)
X	The independent variable—also known as the <i>treatment, factor, or program</i>
Analyze	Analyze qualitative and quantitative data
Feedback	Based on the analyses, researcher makes necessary adjustments, feeds findings back into program
Evaluate	After feedback adjustments, program is evaluated, tested again

Diagram 20.1 Posttest With a Historical Control for the Action Research Approach



Note: The initial test can include data from a cohort control. This will allow the research to compare the results from the treatment group to a previous cohort that was not exposed to the treatment or program.

Diagram 20.2 Pretest and Posttest Design for the Action Research Approach

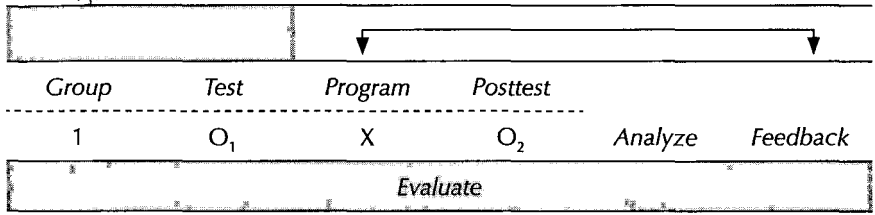


Diagram 20.3 Posttest-Only Design for the Action Research Approach

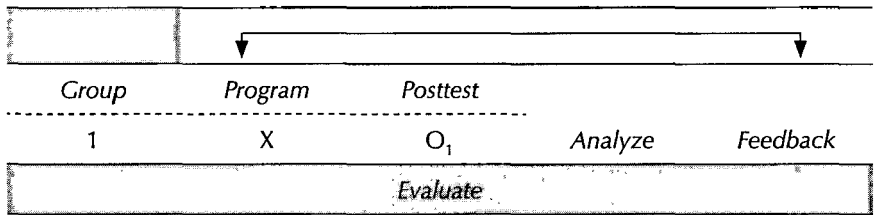
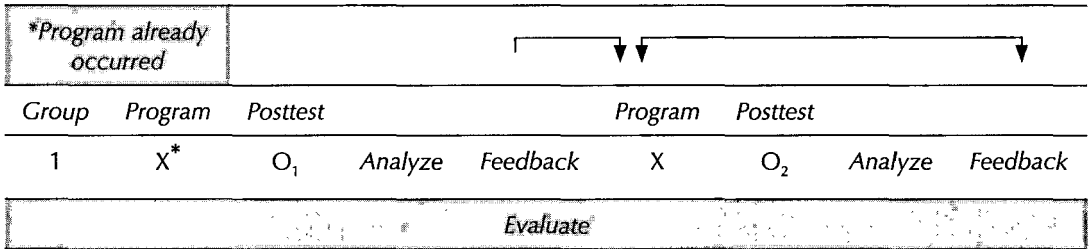


Diagram 20.4 Ex Post Facto Design for the Action Research Approach



Note: The original program already occurred. The data are collected, analyzed, and then fed back to the program; they are reapplied in real time, and then data are collected again, analyzed, and fed back.

Reviewing the Content and Testing Your Knowledge

Exercise

Develop a hypothetical research scenario that would necessitate the use of the **Action Research Approach** and a **Posttest With a Historical Control Design**. The research will be considered nonexperimental.

1. Identify the research scenario, including the general area of focus.
2. Identify the most appropriate approach and then design. Provide a rationale as to why this approach and design would be most appropriate.
3. Develop the appropriate primary research question to be associated with this design.
4. Discuss the sampling strategy and technique to be used.
5. Based on the design, briefly discuss the data collection procedures to be used. Be sure to include the area of focus and targeted sample as part of these procedures.
6. Discuss the themes, theory, and/or phenomenon that would be anticipated to emerge as a result of the examination.
7. Briefly discuss the strengths and limitations associated with this approach and the specific design.

CHAPTER 21

CONCLUSION

◆ PHILOSOPHICAL TENETS

The goal of this guide is to provide practical applications and visual representations of the most common research designs in the fields of education, health and the social and behavioral sciences. We hope that the presentation of each design and the relevant applied examples will encourage researchers to apply many of these theoretically sound designs, which, in our opinion, are underused (particularly mixed methods). This is an applied text, focused on presenting visual aids and real-world examples, to illustrate the key points rather than covering foundational and theoretical issues. However, the importance and relevance of the theory and philosophy related to the various research methods should be noted. That is, there are many theoretical tenets and philosophical principles undergirding the use of a particular method and associated design. More specifically, quantitative researchers focus on testing an a priori theory with an emphasis on deductive reasoning, and they are more in line with postpositivism. Alternatively, qualitatively oriented researchers often use inductive reasoning (or abductive), which is reflective of constructivism. Mixed methods can be viewed more as pragmatic in that this form most efficiently combines both the philosophical approaches of inductive and deductive reasoning. Some researchers would argue that all research should include mixed methods in that the form addresses the complexity of current research problems and counteracts the limitations inherent in using only one type of method. The reader is referred to Creswell (2012)

for an in-depth overview of philosophical approaches to quantitative, qualitative, and mixed methods.

EVALUATION APPROACHES ♦

Within this book, we cover some of the most common research designs in quantitative, qualitative, and mixed methods. We have not included evaluation approaches or program evaluation models within the book. Evaluation approaches are primarily used to judge (or evaluate) the merit or worth of an entire program or the product or processes of a program. Although many evaluation approaches have emerged from the traditional framework of social science research, there is a point when evaluation and research diverge in several key areas. More specifically, the primary goal of research (quasi-experimental or experimental and nonexperimental) is to (a) expand, confirm, or develop theories; (b) seek outcomes; (c) generalize the findings (to the subject or population of interest in quantitative methods); and (d) disseminate the results. Alternatively, the primary goal of an evaluation is to draw judgments based on the findings; however, instead of disseminating the findings, the results are fed back to the stakeholders and ultimately integrated into the program of interest. Another key distinction between research and evaluation is that a researcher develops the research objectives or questions, whereas the stakeholders typically develop the aims or objectives for the evaluator to pursue.

Despite these differences, there are many instances where research and evaluation do overlap (i.e., converge). Based on the objectives set forth by the stakeholders and considering the type of program evaluation model to be employed, the appropriate research design should be embedded within the evaluation approach. Specifically, the process of selecting a research design within a program evaluation can take place once the research questions or objectives have been determined by the stakeholder. The most appropriate research design is then incorporated to answer the stated questions. Logistically speaking, it is usually not feasible or relevant to use experimental research within a program evaluation; however, observational, survey, time-series approaches, or *regression point-displacement* designs (RPD; see Linden, Trochim, & Adams, 2006; Trochim & Campbell, 1996) can serve as strong design alternatives. Most leaders in the field of evaluation agree that mixed methods is the best method to be used by evaluators. Creswell and Plano Clark's (2011) mixed method *multiphase design* is an ideal variant to be combined with program evaluation models.

The reader is referred to Stufflebeam and Coryn (2014) for an in-depth review of evaluation models and applications. To see an exhaustive list of checklists related to evaluation, including the CIPP (context, input, process, product) model, the reader is referred to D. Stufflebeam's Evaluation Checklist at <http://www.wmich.edu/evaluation/checklists>.

♦ DATA ANALYTICS

We want to emphasize again that referring to the research designs presented in this reference guide as “common” can be misleading and does not mean that the designs are less powerful or that the results yielded will have less meaning. Practical and statistical significance can be ensured as long as validity is secured throughout the process (i.e., instrumentation, data collection, analysis, and reporting) and an adequate number of data points and participants are included (i.e., statistical power). Issues related to statistical power and determining the number of participants to include in any given study can be reviewed in Kraemer and Blasey (2015) and a freeware program called G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). The statistical or data-analytic techniques are driven by (a) the research questions or hypotheses and (b) the research design of choice. After the presentation of each design, we recommend the most appropriate statistical procedure (parametric) be used, and we offer recommendations of data-analytic software for qualitative methods. Statistical procedures will vary with the application of each design, and there may be instances when nonparametric procedures should be applied. We refer the reader to Green and Salkind (2013) and Leech, Barrett, and Morgan (2014) for sound texts that detail statistical procedures and techniques using statistical software packages (SPSS), as well as statistical applications using Microsoft Excel 07–13 (e.g., Pace, 2011). We also refer the reader to Bazeley and Jackson (2013) for techniques related to qualitative data analyses.

♦ FINAL REMARKS

There are many different types of approaches to research (some considered to be more obscure) that are not research or method specific, and they are

not discussed in this book. These include collaborative approaches such as (a) systematic review approaches such as quantitative meta-analysis, qualitative metasynthesis, and metastudy; (b) arts-based approaches such as autoethnography, portraiture, and life history (e.g., Butler-Kisber, 2010); and (c) Delphi techniques (see Guyz, Dickson-Swift, Kenny, & Threlkeld, 2015, for discussions related to issues and new approaches to applying the Delphi technique).

These approaches can be applied with quantitative, qualitative, or mixed methods. Regardless of the type of approach, method, or research, one should still adhere to the appropriate tenets of scientific inquiry when examining phenomena in education and the social and behavioral sciences. Although not all types of research are aimed specifically at establishing cause and effect links, at some level all researchers should consider Cook and Campbell's (1979) three conditions for establishing cause and effect: (a) covariation (the change in the cause must be related to the effect), (b) temporal precedence (the cause must precede the effect), and (c) no plausible alternative explanations (the cause must be the only explanation for the effect). Implementing sound research designs is one of the primary steps in controlling for the issues related to plausible alternative explanations and satisfying the required conditions. Cook and Campbell mentioned that their research designs should not be used as templates, but rather as guides to initiate inquiry.

However, based on many years of research and substantiation of the designs presented herein, these are some of the strongest designs applied within education and the social and behavioral sciences. Therefore, researchers can use and apply these designs as presented, with no modifications. Furthermore, each design can be modified to suit the primary stated research question. For example, a series of posttest observations can be added to the basic pretest and posttest control group design as a means to include a time-series component (see designs found in the repeated-measures approach for examples), or a researcher can integrate and combine various methodological and design components in the application of qualitative methods (as discussed with the generic design). Nonetheless, decisions such as these should be based on theoretical tenets and logistical considerations, and we stress that researchers use the most appropriate and parsimonious research design to answer the stated research questions. We also emphasize that researchers employ clarity and consistency when discussing the research design in written research reports and manuscripts submitted for publication. Consistency in terminology and clear descriptions of the design provide the reader with the necessary insight and understanding of the examination at hand.

Appendix A

Less Common Designs for Experimental and Quasi-Experimental Research

The designs presented here in Appendix A (Examples 1–9) are less commonly used but can be useful and appropriate under the correct circumstances. Some of these designs can be considered *pattern matching designs* (i.e., combining various design features into one design to improve the overall internal validity). These designs are structured for a variety of research scenarios when it is not logistically possible to use random assignment. The addition of various types of comparison groups and the addition of multiple pretest or posttest measures strengthens these research designs (in terms of internal validity) for various applications of quasi-experimental research. These quasi-experimental research designs are particularly ideal for researchers conducting examinations in the educational sector, considering that random assignment is rarely feasible. Examples A5, A10, and A11 are designs for experimental research rarely applied in the social sciences, but they are considered strong designs.

Example A1 Proxy Pretest–Posttest Design

<i>Group</i>	<i>Assignment</i>	<i>Proxy Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
1	NR	O_{A1}	X	O_{B1}
2	NR	O_{A1}	—	O_{B1}

Time ►

Note: This is an example of a between-subjects approach with a proxy pretest and posttest control group design. The proxy pretest allows the researcher to compare the “treatment effects” of O_{B1} to a proxy variable (O_{A1}). This design is useful when a program or intervention has already started, and the researcher was not able to collect the same pretest measure that is being collected for the posttest. Therefore, data (archived) from a proxy variable can be collected that is considered conceptually similar to the posttest and can closely estimate pretest performance. For example, if a reading intervention is being implemented and a reading achievement test is being collected for the posttest, then a possible proxy variable would be the students’ GPA prior to the intervention. Although the proxy pretest provides a measure of control, selection bias remains a major threat to the internal validity of this design.

Example A2 Double Pretest and Posttest Control Group Design

<i>Group</i>	<i>Assignment</i>	<i>Pretest</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
1	NR	O_1	O_2	X	O_3
2	NR	O_1	O_2	—	O_3

Time ►

Note: This is an example of a between-subjects approach with a double pretest and posttest control group design. The double pretest allows the researcher to compare the treatment effects between O_1 to O_2 and then from O_2 to O_3 . A major threat to internal validity with this design is testing, but it controls for selection bias and maturation.

Example A3 Pretest and Posttest Historical Control Group With a Pretest and Posttest Design

<i>Group</i>	<i>Assignment</i>	<i>Pretest</i>	<i>Posttest</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
1	NR	O ₁	O ₂	—	—	—
2	NR	—	—	O ₁	X	O ₂

Time ►

Note: This is an example of a between-subjects approach with a pretest and posttest historical control group and pretest and posttest design. The historical control group allows the researcher to compare the treatment effects between O₁ to O₂ from the historical control to O₁ to O₂ of the treatment group. History and selection bias are the two most prevalent threats to the internal validity of this design.

Example A4 Posttest Only With a Historical Control Group Design

<i>Group</i>	<i>Assignment</i>	<i>Test</i>	<i>Treatment</i>	<i>Posttest</i>
1	NR	O ₁	—	—
2	NR	—	X	O ₁

Time ►

Note: This design would be designated as quasi-experimental research using a between-subjects approach with a posttest only and a historical control group design. History and selection bias would be the biggest threats to internal validity.

Example A5 Pretest and Posttest Control Group Plus Historical Control Group Design

<i>Group</i>	<i>Assignment</i>	<i>Test</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
1	R	—	O ₁	X	O ₂
2	R	—	O ₁	—	O ₂
3	NR	O ₁	—	—	—

Time ►

Note: This is an example of a between-subjects approach pretest and posttest control group design with the addition of a historical control group. The pretest and posttest aspect can be applied with or without random assignment. The historical control is typically a cohort control and helps to control for testing effects, which are a major threat to internal validity in designs that include pretest measures.

Example A6 Regression Point-Displacement (RPD) Design

<i>Unit</i>	<i>Assignment</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
1	NR	O_1	X	O_2
2*	NR	O_1	—	O_2

Time ►

Note: This is an example of a between-subjects approach with a pretest and posttest control group RPD design. This design is best used for program evaluations or community-based research. It is difficult to infer causation in community-based examinations based on the evaluation of a single unit (or community) using the basic one-group pretest and posttest design. Therefore, for the comparison unit (2*), data can be collected from a heterogeneous set of units (or communities) and then collapsed and compared to the single unit that received the treatment. If logistically feasible, a time-series component can be added with a series of multiple pretests and posttests. A form of regression analysis is used to analyze the results. We refer the reader to Trochim and Campbell (1996) and Linden, Trochim, and Adams (2006) for further explanations on the RPD design and the most common threats to the internal validity of this design.

Example A7 One-Group Double Pretest and Posttest Design

<i>Group</i>	<i>Assignment</i>	<i>Pretest</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
1	NR	O_1	O_2	X	O_3

Time ►

Note: This is an example of a one-group within-subjects approach with a double pretest and posttest design. The double pretest allows the researcher to compare the treatment effects between O_1 to O_2 and then from O_2 to O_3 . A major threat to internal validity with this design is history, but it controls for testing and maturation. The one-group design is not considered as strong as the two-group variant of this design.

Example A8 One-Group Treatment-Removed Design

<i>Group</i>	<i>Assignment</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Midtest</i>	<i>Midtest</i>	<i>Treatment</i>	<i>Posttest</i>
1	NR	O_1	X	O_2	<i>Time Delay</i>	O_3	— O_4

Time ►

Note. This is an example of a one-group pretest multiple posttest with the treatment-removed design. The goal of this design is to establish the change in the outcome based on the presence or absence of the treatment. Therefore, the researcher would assess the change from O_1 to O_2 and compare that to the change from O_3 to O_4 , hypothesizing that in the absence of the treatment, the outcome would move in the opposite direction compared to that when it is present. It is assumed that the effects of the treatment should be expected to dissipate over time. Due to the lack of a comparison group, a variable (confounding) not controlled for can account for the change in the outcome; therefore, a large sample is required in order to minimize the negative impact on the statistical conclusion validity.

Example A9 One-Group Repeated-Treatment Design

Group	Assignment	Pretest	Treatment	Posttest	Treatment	Midtest	Treatment	Posttest
1	NR	O ₁	X	O ₂	—	O ₃	X	O ₄

Time ►

Note: This is an example of a one-group pretest and posttest design repeated over time. The aim of this design is to verify that the change from O₁ to O₂ is similar in change from O₃ to O₄, hypothesizing that in the absence of the treatment, the outcome will move in the opposite direction between O₂ and O₃. The effects of the treatment implemented should be expected to dissipate over time, and the researcher should include a considerable delay between the initial treatment and the second application. Due to the lack of a comparison group, a variable not controlled for can account for the change in the outcome; therefore, a large sample is required in order to minimize the negative impact on the statistical conclusion validity. Testing, maturation, and history are major threats to the internal validity of this design.

Example A10 3-Factor Crossover Design

Group	Assignment	Pretest	Treatment	Midtest	Treatment	Midtest	Treatment	Posttest
1	R	O ₁	X _A	O ₂	X _B	O ₃	X _C	O ₄
2	R	O ₁	X _B	O ₂	X _A	O ₃	X _C	O ₄
3	R	O ₁	X _C	O ₂	X _A	O ₃	X _B	O ₄

Time ►

Note: This is an example of a repeated-measures approach 3-factor crossover design. Each group serves in one condition, and the conditions are counterbalanced to control for sequencing effects. This design can be modified in multiple ways, such as adding additional factors, introducing the same factor more than once in each condition, and including more observations. The variation of treatment orders can go up to 12 while including one participant per condition ($N = 12$). Based on this variation, the participant serves as his or her own control, which is an intended feature built into designs for repeated-measures approaches.

Example A11 3 × 3 Graeco-Latin Square Design

Blocking Factor (Levels 1–3)	Blocking Factor (Levels 1–3)		
	1	2	3
1	A_α	B_γ	C_β
2	B_β	C_α	A_γ
3	C_γ	A_β	B_α

Note: This is an example of a 3 × 3 Graeco-Latin square design. Similar to the Latin-square design, this design is a 1-factor model, but instead of two blocking factors, it includes a third extraneous factor and is denoted as α , γ , and β . As with the Latin-square design, this design is best suited for research in agriculture and engineering; few scenarios warrant the use of such a design within the social sciences. In addition, with the use of human subjects, sequencing effects are a major threat to the internal validity of this design application. The analysis of means (ANOM) is the appropriate analysis for this design.

RESEARCHER CHALLENGE ♦

We present a complex design that has yet to be implemented in the social and behavioral sciences. We understand that few scenarios warrant the application of such a design. However, if anyone would like to take this challenge on, we would like to know and be part of the peer review committee once the manuscript is submitted for publication. Good luck!

Example A12 3 × 3 Repeated Latin-Square Design

		Blocking Factor 2			Blocking Factor 2			Blocking Factor 2		
Blocking Group	Factor 1	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
1	Level 1	A	B	C	C	A	B	B	C	A
2	Level 2	B	C	A	A	B	C	C	A	B
3	Level 3	C	A	B	B	C	A	A	B	C

Note: An important design assumption for the Latin-square design is that there is no interaction between the blocking factors, as well as the primary factor. However, all combinations of each level of the three factors (2 blocking factors and 1 primary factor) are being tested with this repeated design; therefore, the interaction effects become a primary basis for inference and do not contaminate the results of the main effects. The appropriate statistic would be the three-way between-subject ANOVA. Statistical power: The number of levels of the two blocking factors is multiplied to determine the number of runs or number of participants required for each condition.

Example Research Question: What are the effects of different styles of putting on performance when controlling for varying levels of anxiety and putt difficulty?

Null Hypothesis

Main effects. There is no difference in performance between the different levels of putt difficulty. There is no difference in performance between the different levels of anxiety. There is no difference in performance between the different putting styles.

Interaction effects. There is no interaction between anxiety and putt difficulty on performance. There is no interaction between anxiety and putting style on performance. There is no interaction between putt difficulty and putting style on performance. There is no interaction between putt difficulty, putting style, and anxiety on performance.

$k = 3$ factors (2 blocking factors and 1 primary factor)

L_1 (Block) = Anxiety (3 levels)—High, Medium, Low

L_2 (Block) = Putt Difficulty (3 levels)—Easy, Medium, Hard

L_3 (Primary) = Putt Style (3 levels)—Long, Belly, Standard

$n = 9$ subjects per condition (number of levels of the two blocking factors are multiplied to determine the number of runs or number of participants per condition— $3_1 * 3_2 = 9$)

		Putt Difficulty			Putt Difficulty			Putt Difficulty		
Group	Anxiety	Easy	Medium	Hard	Easy	Medium	Hard	Easy	Medium	Hard
1 ($n = 9$)	High	A	B	C	C	A	B	B	C	A
2 ($n = 9$)	Medium	B	C	A	A	B	C	C	A	B
3 ($n = 9$)	Low	C	A	B	B	C	A	A	B	C
		Day 1			Day 2			Day 3		

Note: A = Long; B = Belly; and C = Standard. This design allows for all subjects to be exposed to every combination of the level of factors, which could also be achieved by a $3 \times 3 \times 3$ factorial design. However, the order of conditions is set up horizontally instead of vertically, which allows for the three groups ($N = 27$) to remain intact over time, as opposed to including nine different groups for a between-subjects $3 \times 3 \times 3$ design. This would require a total of 81 participants. Although the conditions are counterbalanced, sequencing effects are still a threat to internal validity; therefore, based on theoretical and logistical considerations, a certain amount of time should elapse between conditions (i.e., a “wash out” period) to minimize the impact of sequencing effects. Another option is to counterbalance the order of the groups; so, for example, Group 1 would be exposed to high anxiety on Day 1, low anxiety on Day 2, and medium anxiety on Day 3. Treatment fidelity is paramount in this particular design.

Appendix B

Types of Case Study Designs

Case Study Designs: Creswell (2014)

Creswell placed the case study within the ethnographic approach; however, the case study can be applied within the framework of any of the approaches detailed under qualitative methods.

<i>Type</i>	<i>Definition</i>
Intrinsic	The examination of a unique case
Instrumental	The examination of a case to provide insight into an issue or specific theme
Multiple Instrumental (also known as <i>Collective Case Study</i>)	The examination approach, which is the same as instrumental, but with multiple cases

Case Study Designs: Yin (2013)

According to Yin (2013), *holistic* refers to identifying and collecting data from a single unit of analysis, whereas *embedded* refers to collecting data from multiple units of analysis. Yin also indicated that his case study designs can be categorized and conducted as exploratory, descriptive, or explanatory (causal) investigations.

<i>Type</i>	<i>Definition</i>
Single Case–Holistic	
Critical case	Examine a well-formulated theory.
Unique case	Examine an extreme case.
Representative case	Examine a typical or average case.
Revelatory case	Examine a phenomenon that was previously inaccessible.
Longitudinal case	Examine the same cover over a period of time.
Single Case–Embedded	
Critical case	Examine a well-formulated theory.
Unique case	Examine an extreme case.
Representative case	Examine a typical or average case.
Revelatory case	Examine a phenomenon that was previously inaccessible.
Longitudinal case	Examine the same cover over a period of time.
Multiple Case–Holistic	
Literal replication	Select and examine each case so that all cases are presumed to predict similar results.
Theoretical replication	Select and examine each case so that all cases are presumed to predict contrasting results, but for anticipatable reasons.
Multiple Case–Embedded	
Literal replication	Select and examine each case so that all cases are presumed to predict similar results.
Theoretical replication	Select and examine each case so that all cases are presumed to predict contrasting results, but for anticipatable reasons.

Appendix C

Five Types of Mixed Methods Designs

<i>Families of Mixed Methods Designs (according to Tashakkori & Teddlie, 2010a)</i>	
<i>Design</i>	<i>Procedures</i>
Parallel mixed	<ul style="list-style-type: none"> • Mixing occurs in a parallel manner. • Data are collected simultaneously (or with some time lapse). • QUAL and QUAN phases answer related aspects of the same research questions.
Sequential mixed	<ul style="list-style-type: none"> • Mixing occurs across chronological phases (QUAL, QUAN). • Questions or procedures from one method emerge from, or depend on, the one prior. • Research questions are related to one another and may evolve.
Conversion mixed	<ul style="list-style-type: none"> • Parallel design is used. • Mixing occurs when one type of data is transformed and analyzed both qualitatively and quantitatively. • This is used to answer related aspects of the same research questions.
Multilevel mixed	<ul style="list-style-type: none"> • Parallel or sequential design is used. • Mixing occurs across multiple levels of analysis. • QUAN and QUAL data from these different levels are analyzed and integrated to answer aspects of the same (or related) research questions.
Fully integrated mixed	<ul style="list-style-type: none"> • Mixing occurs in an interactive manner at all stages of the study. • At each stage, one approach affects the formulation of the other. • Multiple types of implementation processes occur.

Note: Tashakkori & Teddlie (2010a) also present a quasi-mixed design (monostrand conversion design), in which the “mixed” aspect refers to the *quantitizing* or *qualitizing* of data. In other words, the researcher would convert (“mix”) one form of data (QUAL) to another form (QUAN) and only use the converted form of the data (QUAN or QUAL) to answer the research questions.

Appendix D

Reporting Preliminary Statistical Findings

In this appendix, we provide clear guidelines (including examples) that demonstrate how to write preliminary results using APA style. This appendix includes elements related to many statistical tests. Both the use of narrative (words) and numbers are strung together to provide concise examples of how to address each aspect of the results section. Throughout this appendix, data from published results of various research articles are provided.

A following brief decision tree is followed by the quick-reference image of a normal distribution and then three sections. The first section covers the preliminary analyses required across many types of research. For example, effect size, power, reliability and validity, and manipulation checks. The second section covers basic descriptive analyses often used prior to the primary analyses, including simple correlational analyses. The third section demonstrates how to present some data visually in proper APA-formatted tables and figures. For those interested in a more thorough discussion related to dealing with missing data, see Groenwold, Donders, Roes, Harrell, and Moons (2012).

STATISTICAL DECISION TREE ♦

Statistical Tests for One Interval IV

<i>Independent Variable</i>	<i>Dependent Variable</i>	<i>Statistical Test</i>
One Interval IV	Interval	1) Correlation 2) Simple linear regression
	Ordinal	2) Nonparametric correlation
	Nominal	3) Simple logistic regression

Statistical Tests for One or More Interval IV and/or One or More Categorical IV

<i>Independent Variable</i>	<i>Dependent Variable</i>	<i>Statistical Test</i>
One or more Interval IV and/or one or more Categorical IV	Interval	1) Multiple regression 2) Analysis of covariance
	Categorical	3) Multiple logistic regression 4) Discriminant analysis

INFERENCEAL ♦

Statistical Tests for Zero Independent Variables

<i>Independent Variable</i>	<i>Dependent Variable</i>	<i>Statistical Test</i>
Zero IVs (One sample/group)	Interval and existing value	1) One-sample <i>t</i> test
	Ordinal and existing value	2) One-sample median
	Nominal (two categories) and existing value	3) Binomial test
	Nominal and existing value	4) Chi-square goodness of fit

Statistical Tests for One Independent Variable

<i>Independent Variable</i>	<i>Dependent Variable</i>	<i>Statistical Test</i>
One IV with two levels (Independent groups)	Interval	1) Independent sample t test
	Ordinal	2) Wilcoxon-Mann Whitney
	Nominal (2 categories)	3.1) Chi-square 3.2) Fisher's exact

<i>Independent Variable</i>	<i>Dependent Variable</i>	<i>Statistical Test</i>
One IV with two or more levels (Independent groups)	Interval	1) One-way ANOVA
	Ordinal	2) Kruskal Wallis
	Categorical	3) Chi-square

<i>Independent Variable</i>	<i>Dependent Variable</i>	<i>Statistical Test</i>
One IV with two levels (Dependent/paired groups)	Interval	1) Paired samples t test
	Ordinal	2) Wilcoxon signed ranks
	Categorical	3) McNemar

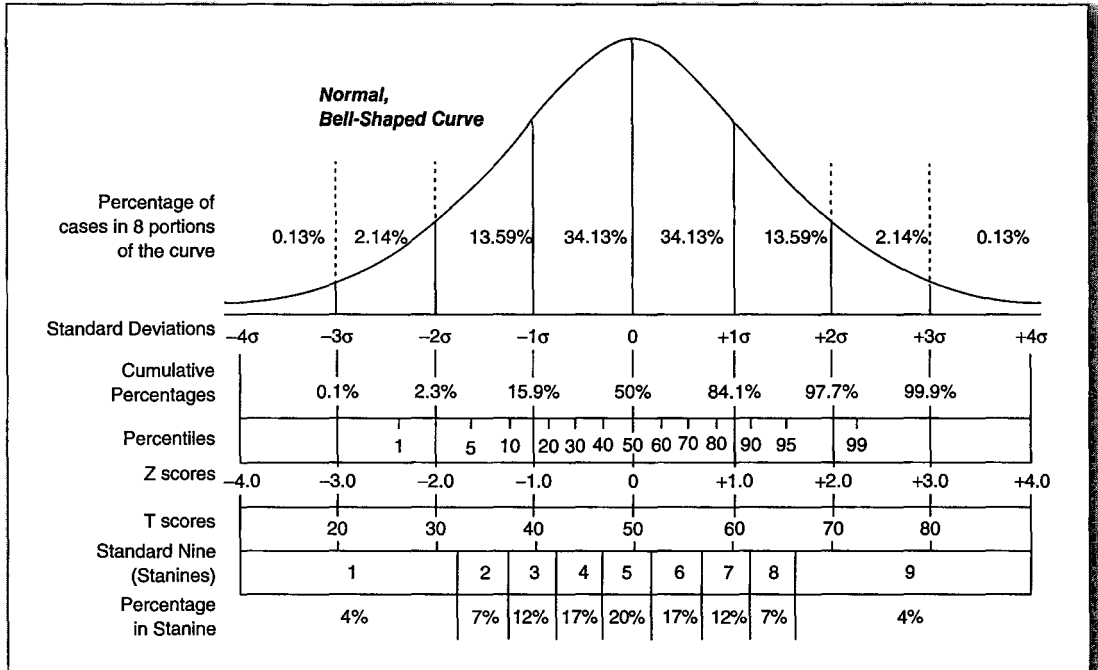
<i>Independent Variable</i>	<i>Dependent Variable</i>	<i>Statistical Test</i>
One IV with two or more levels (Dependent/paired groups)	Interval	1) One-way repeated measures ANOVA
	Ordinal	2) Friedman test
	Categorical	3) Logistic regression

Statistical Tests for Two or More Independent Variables

<i>Independent Variable</i>	<i>Dependent Variable</i>	<i>Statistical Test</i>
2 or more IVs (Independent groups)	Interval	1) 2- or 3-way ANOVA
	Categorical	2) Logistic regression

Statistical Tests for One Independent Variable With Two or More Levels

Independent Variable	Dependent Variable	Statistical Test
One IV with two or more levels (independent group)	Interval	1) One-way MANOVA



Note: The bell-shaped curve represents a "normal" distribution and is a key characteristic to the probability theory associated with parametric statistics. The reader interested in further understanding this in relation to statistics should learn about the *central limit theorem* and the *law of large numbers*.

SECTION 1: PRELIMINARY ♦

1) Manipulation Check

A manipulation check allows the researchers to check if the independent variable(s) (variable(s) that are manipulated) work the way they are intended. In other words, it is a procedure to test if the levels of the

independent variable differ on the dependent variable. It is similar to the concept of fidelity.

For example, Hofer, Burkhard, and Allemand (2015) examined if the stimulus they were using to elicit emotional responses worked as intended:

To examine whether the emotional stimulus was successful in eliciting negative emotions and whether younger and older adults differed in their emotional reactions, we conducted a 2 (age group: young vs. old) · 2 (time: T1 vs. T2) mixed ANOVA for each emotional reaction. (p. 51)

The result of their analyses suggested that the IV was working as intended (see the following).

Table D1 Emotions as a Function of Age

	<i>Age Group</i>			
	<i>Younger Adults (18–28)</i>		<i>Older Adults (62–87)</i>	
	M	SD	M	SD
<i>Sadness</i>				
T1	1.66Aa	0.90	1.38Ab	0.68
T2	3.41Ba	1.06	3.10Bb	1.19
<i>Anger</i>				
T1	1.37Aa	0.84	1.19Aa	0.59
T2	3.43Ba	1.15	3.30Ba	1.30

Source: Hofer et al., 2015, p. 51.

The following from Reynolds, McCambridge, Bissett, and Considine (2014) is an excellent format to present your manipulation check:

Analyses began by assessing whether disgust was successfully induced using a 2 (disgust vs. control condition) x 2 (start vs. end of session) MANOVA. . . . A significant difference in state disgust was found across conditions, Wilks' $\Lambda = .90$, $F(2, 77) = 4.40$, $p = .016$, $\eta^2p = .10$; the experimental group reported more disgust than the

control group at both the beginning, $F(1, 78) = 5.65, p = .020, \eta^2 p = .07$, and end of the session, $F(1, 78) = 7.75, p = .007, \eta^2 p = .09$. The absence of a difference between state disgust at the beginning and end of the session showed that disgust was sustained throughout the experiment, Wilks' $\Lambda = 1.00, F(1, 78) = .11, p = .739, \eta^2 p = .001$. Importantly, in terms of eliminating other possible affective bases for avoidance, the induction was specific to disgust, as embarrassment, $F(2, 77) = .12, p = .884, \eta^2 p = .003$, and fear, $F(2, 77) = 1.20, p = .308, \eta^2 p = .03$, did not vary as a function of condition.

Given that participants in both control and experimental conditions were presented with numerous stimuli designed to elicit disgust over and above the “olfactant” (e.g., both groups were presented with an apparently “used” stoma bag), the subsequent analysis of state disgust concentrates on DES disgust scores at the start of the laboratory session—that is, before these other factors had been introduced to participants. (pp. 1498–1499)

2) Effect Size

Examining the magnitude, or size, of an effect

Standardized Mean Differences (for Comparing Differences Between Treatment Means)

Cohen's d

This is a standardized mean effect—the mean difference between two groups in standard deviation units. The meaning of effect size may vary by context, but Cohen (1988) offers the following general rule of thumb:

.8 = large (8/10 of a standard deviation unit)

.5 = moderate (1/2 of a standard deviation)

.2 = small (1/5 of a standard deviation)

Example From the Literature

“These findings were associated with a moderate effect size across multiple measures (Cohen's $d = 0.56$ – 0.58), providing strong evidence for an impact of melancholia on vagally mediated, cardiac function” (Kemp, Quintana, Quinn, Hopkinson, & Harris, 2014, p. 5).

Hedges's g

Hedges's g incorporates an adjustment that removes the bias of Cohen's d .

Example From the Literature

"These effects remained stable at follow-up. Moderator analyses revealed cognitive-behavioral treatment to be significantly better than other psychological treatments in short-term pain reduction (Hedges's $g = 0.60$, 95% CI: 0.46–0.76)" (Glombiewski et al., 2010, p. 280).

Glass's Δ

Glass's delta is the mean difference between the experimental and control group divided by the standard deviation of the control group.

Example From the Literature

"The pre- to post-treatment effect sizes (Glass's delta) were 1.24 for the University of California at Los Angeles Post-Traumatic Stress Disorder Reaction Index and 1.96 for the Children's Global Assessment Scale" (Kameoka et al., 2015, p. 1).

*Measures of Association
(Correlation and Explained Variance)**Pearson's r*

A standardized measure of the strength and direction of a linear relationship between two variables ranging from -1 for a perfect negative relationship and 1 for a perfect positive relationship.

Example From the Literature

"Students increased their knowledge of HIV/AIDS (Pearson's measure of effect size $r = 0.74$) and the risk of acquiring HIV infection ($r = 0.68$) statistically significantly ($p = 0.001$). Girls (risk estimation: $r = 0.78$, knowledge: $r = 0.81$) improved much more than boys (risk estimation: $r = 0.57$, knowledge: $r = 0.62$)" (Hlavinkova, Mentel, Kollarova, & Kristufkova, 2014, p. 905).

Spearman's r_s

Used when both X and Y are measured on a ranked scale.

Example From the Literature

“The relationship between the number of teamwork principles taught and effect size achieved a Spearman’s correlation of .74 ($p = .01$) for overall effect size and .64 ($p = .03$) for median skills/ behaviors effect size” (Chakraborti, Boonyasai, Wright, & Kern, 2008, p. 846).

Point biserial r_{pb}

Used when X is dichotomous and Y is continuous.

Example From the Literature

The outcome variable was dichotomous, which necessitated the use of a point-biserial correlational analysis (r_{pb}). The tests of significance were set at the .05 level. Out of the behavioral variables, two demonstrated the strongest significant correlation with offender status: (a) MACI-Eating Dysfunction $r(93) = .19, p < .05$, and (b) MACI-Delinquent Predisposition $r(93) = .19, p < .05$. (Kennedy, Burnett, & Edmonds, 2011, p.320)

Phi coefficient (φ)

Used when both X and Y are dichotomous (i.e., both variables and both outcomes can be arranged on a 2×2 contingency table).

Example From the Literature

“The reliability of end-of-rotation evaluations was adequate (fellows, phi coefficient [φ] = 0.68; faculty [including programme directors], $\varphi = 0.71$)” (Park, Riddle, & Tekian, 2014, p. 614).

3) Reliability

Inter-Item Reliability (Cronbach’s Alpha)

Cronbach’s alpha is a measure of internal consistency and should be included with the description of the measure (e.g., measures of central tendency and dispersions).

Example From the Literature

To validate test-retest reliability, 32 patients with iRBD were tested again at an interval of 2–4 weeks. Cronbach’s alpha was computed

to assess the internal consistency of the RBDSQ-K. The criterion value of Cronbach's alpha was more than 0.70 for item homogeneity. Intraclass correlation coefficient for RBDSQ-K total score was also computed and each item of the RBDSQ-K was assessed with Cohen's kappa coefficient. Item-test correlations were calculated to measure the relation between the response to the specific test questions and the RBDSQ-K total score. (Lee, Paek, & Ryu, 2015, in press)

Interrater Reliability (Cohen's Kappa and Scott's Pi)

Cohen's kappa and Scott's pi measure the internal consistency of raters applying some scale of measurement.

Examples From the Literature

The Cohen's kappa coefficient was substantial (≥ 0.61) for RBDSQ-K Items 7 and 10 (Table 3). The strength of the correlation between each individual item of the RBDSQ-K and the total RBDSQ-K score ranged from 0.249 to 0.749 (all $p < 0.001$).

A purposive sample of 50 Facebook profiles created by environmental advocacy groups was content analyzed by two trained coders using an 82-item questionnaire. Intercoder reliability scores were calculated using Scott's pi (Holsti, 1969) for dialogic strategies and Holsti's coefficient of reliability (Holsti, 1969) for dialogic outcomes. Scores ranged from 61% to 87% with Scott's pi and 90% to 100% with Holsti's. (Bortree & Seltzer, 2009, p. 317)

4) Validation

Validation is the systematic process of demonstrating that an instrument is valid (the extent to which scores generated by an instrument measure the variable they are intended to measure). The following are examples of some ways to validate four common forms of validity seen throughout the literature. It should be noted that the concept of validity is often considered unitary, and many argue that the following breakdown is an artificial representation of a much more complex interconnected single construct.

1. *Construct validity*—Does the test measure the psychological construct that it claims to measure? This is often measured with correlation of

test scores with other tests that are accepted as measuring that construct or of test scores with direct observations of behavior. Factor analysis is often used for multiscale measure.

2. *Face validity*—Does the test look as if it measures what it is intended? This is usually done with content experts in the field of study, and it is seldom measured statistically.
3. *Content validity*—Is the content of the test valid for measuring what it claims to measure? It is usually assessed by expert judges rather than statistical analysis.
4. *Criterion-related validity*—Do the test scores correlate with an external criterion? This is often used in academics, trainings, and job performance. Correlation is commonly used for validation, although other techniques may be applied (e.g., significance testing for differences between good and poor performers; multiple regression to produce specification equations for selection).

5) Power

The premise of statistical power is the process of determining the number of subjects or how large a sample size should be to ensure that the null hypothesis can be reliably rejected without (or with little chance) of error. There are several factors required that need to be entered into the equation when computing an a priori power analysis. The most difficult item to determine is the critical effect size (denoted as *delta* Δ). The critical effect size is derived from preliminary evidence (sometimes through pilot studies) and provides the parameter that judges how strong the theory must be that would be considered a “contribution” to the field (see Kraemer & Blasey, 2015, for more on this topic). The typical parameters required for estimating power are power, alpha (σ), and critical effect size (Δ). Other forms of power analysis require different parameters.

Example

A psychologist plans to conduct an experiment to examine the effectiveness of a cognitive-behavioral intervention on levels of happiness in retirees. She will employ a two-group pre- and posttest design with a control. Her pilot data suggested that mean levels of happiness for participants who received the intervention were 12.8 points higher with a standard deviation of 1.9 than those who were in the attention-control group.

A power analysis will be performed to determine the sample size estimation for the two-group design. The data from the pilot study revealed scores for the intervention group ($M = 87.8$ $SD = 1.9$) compared to the attention-control group ($M = 75.0$, $SD = 2.3$) were 12.8 points higher for those who received the therapy. The effect size indicated a “medium” effect (partial eta squared was $\eta^2 = .125$). Therefore, the psychologist will set the alpha (α) at .05 (one-tailed), the power at .80, and the critical effect size (Δ) at 0.5, which results in $N = 27$ (see Cohen, 1988, for a presentation of master tables for one- and two-tailed 5% and 1% tests of significance).

◆ SECTION 2: DESCRIPTIVE

1) Frequencies and Percentages

N = Size of the total data set

n = Size of one group or cell

% = Percentages

The participants were recruited from a geriatric center ($N = 100$); men ($n = 40$) represented 40% of the sample, and women ($n = 60$) represented 60%.

Example From the Literature

Ninety-eight women originally served as targets; however, we removed participants who were 31 or older (3 SD s above the mean; $n = 3$) and non-heterosexual participants ($n = 2$). This left us with a final sample of 93 women (age $M = 19.27$, $SD = 1.41$). Their reported ethnicities are as follows: 46% Caucasian, 25% Hispanic, 16% East Asian, 7% Black, 3% South Asian, 1% Middle Eastern, and 2% from other ethnicities. Raters consisted of 115 women and 117 men. (Perilloux, Cloud, & Buss, 2013, p. 491)

2) Measures of Central Tendencies and Dispersions

M = Mean (measure of central tendency)

SD = Standard deviation (measure of dispersion)

Table D2 Descriptive Statistics for Predictor Variables

<i>Independent Variable</i>	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>N</i>
JJSP reading percentile	25.60	20.90	1–97	95
K-BIT vocabulary	76.92	17.69	40–111	95
K-BIT composite	80.57	16.226	48–121	95
PPVT-III	80.64	14.93	58–116	95
WRAT-3 reading subtest	85.57	18.66	45–117	95

Example From the Literature

“We excluded data from non-heterosexual individuals ($n = 14$), leaving a final sample of 113 women and 105 men (age $M = 18.68$, $SD = 2.10$)” (Perilloux, Cloud, & Buss, 2013, p. 491).

3) Correlation*Spearman Rank-Order—What to Report*

df = Degrees of freedom

r_s = Observed, r_s value

p = Significance level

Example From the Literature

“The Spearman rank order correlation coefficient for the 58 cases and two sets of total PCL-R scores was $r = .85$, $p < .001$, two-tailed (that was identical to the ICC for consistency agreement, single measures)” (Harris, Rice, & Comier, 2013, p. 1356).

Pearson Product-Moment—What to Report

df = Degrees of freedom

r = Observed r value

p = Significance level

M = Mean

SD = Standard deviation

Table D3Correlations Among Behavioral Variables and Offender Status: Male and Female
($n = 95$)

Subscale	1	2	3	4	5	6	7
1. Offender status	—	.03	-.18*	-.11	-.07	.15 [†]	.04
2. Social stress		—	.46**	.70**	.49**	.43**	-.67
3. Inhibited			—	.31**	.30**	.42**	-.37**
4. Somatization				—	.47**	.44**	-.58**
5. Substance abuse					—	.29**	-.76**
6. Eating dysfunction						—	-.33**
7. Conforming							—

Note: * $p < .05$. ** $p < .01$. [†] $p = .068$. One-tailed.

Example From the Literature

“The Pearson Product-Moment Correlation test, carried out in the three study groups, showed a significant inverse correlation between QTcD and MMSE score ($r = -0.357$; $p < 0.01$) in the group of MCI patients, only” (Coppola et al., 2013, p. 632).

◆ SECTION 3: TABLES AND FIGURES

1) Tables

The following styles of the tables are in accordance with the *Publication Manual of the American Psychological Association* (APA, 2010). We refer the reader to the *Concise Rules of APA Style* (2010) pocket guide for a quick reference tool for learning how to apply APA style to documents and manuscripts.

Table X Estimated Time (Milliseconds) for Reaction Time in High- and Low-Anxiety Conditions

Condition	M (SD)	95% CI	
		LL	UL
High-anxiety	.87 (.09)	.78	1.03
Low-anxiety	.68 (.10)	.54	.79

Note: CI = Confidence interval; LL = Lower limit; UL = Upper limit.

Table X Estimated Time (Milliseconds) for Reaction Time in High- and Low-Anxiety Conditions for Veterans and Rookies

Condition	Veterans			Rookies		
	n	M (SD)	95% CI	n	M (SD)	95% CI
High-anxiety	11	.81 (.07)	[.69, .89]	14	.90 (.14)	[.72, .94]
Low-anxiety	9	.65 (.11)	[.53, .76]	11	.70 (.13)	[.61, .82]

Note: CI = Confidence interval.

Table X Summary of Intercorrelations, Means, and Standard Deviations for Scores on the QOL, BDI, and BAI as a Function of SES

Measure	1	2	3	M	SD
1. QOL	–	.38*	.64	58.3	8.2
2. BDI	.57*	–	.23	62.8	15.8
3. BAI	.34	.21	–	74.3	16.4
M	54.2	59.6	78.2		
SD	12.2	18.4	21.9		

Note: QOL = Quality of life; BDI = Beck's depression inventory; BAI = Beck's anxiety inventory; SES = Socioeconomic status. Intercorrelations for high SES are presented above the diagonal, and intercorrelations for low SES are presented below the diagonal. Means and standard deviations for low SES are presented in the horizontal columns, and means and standard deviations for high SES are presented in the vertical columns. * $p < .01$.

Table X

Contrast Between Veterans and Rookies for Reaction Time in Low- and High-Anxiety Conditions

Condition	Veterans		Rookies		t(45)	p	95% CI		d
	M	SD	M	SD			LL	UL	
High-anxiety	.81	.07	.90	.14	1.45	.04	.73	.98	.65
Low-anxiety	.65	.11	.70	.13	1.93	.12	.59	.76	.84

Note: CI = Confidence interval; LL = Lower limit; UL = Upper limit. d = Cohen's d.

2) Figures

The following styles of the figures are in accordance with the *Publication Manual of the American Psychological Association* (APA, 2010). The most common types of figures are graphs, charts, maps, drawings, and

Figure D1 Sample Figure Representing Outcome Data

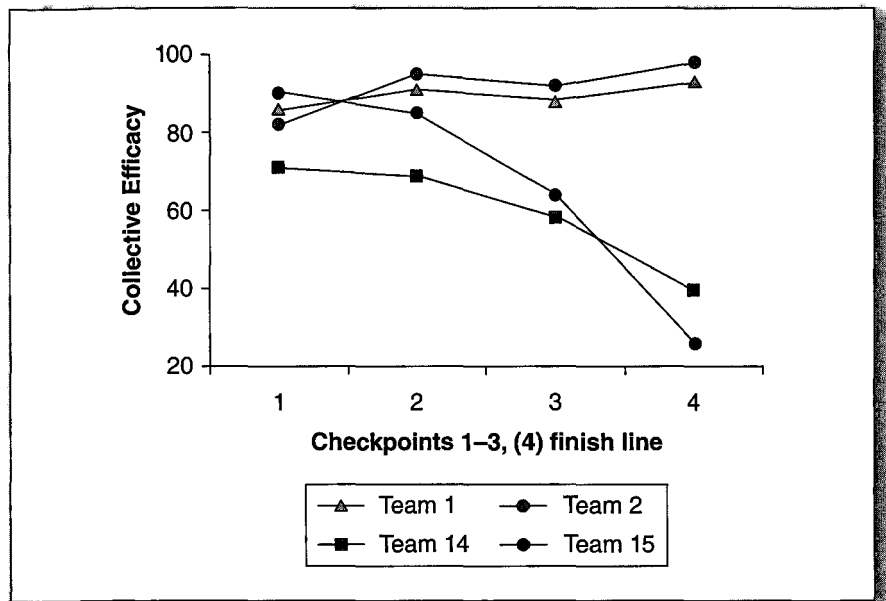


Figure X. Line graph displaying the collective efficacy scores taken throughout the race for teams that finished 1st, 2nd, 14th, and 15th.

photographs. Use figures only when the visual image adds substantively to the meaning of the text or cannot be properly interpreted through a table. Typically, the general rule is if data are only duplicated from either a table or text, then a figure is not necessary.

Figure D2 Sample Figure Representing Outcome Data

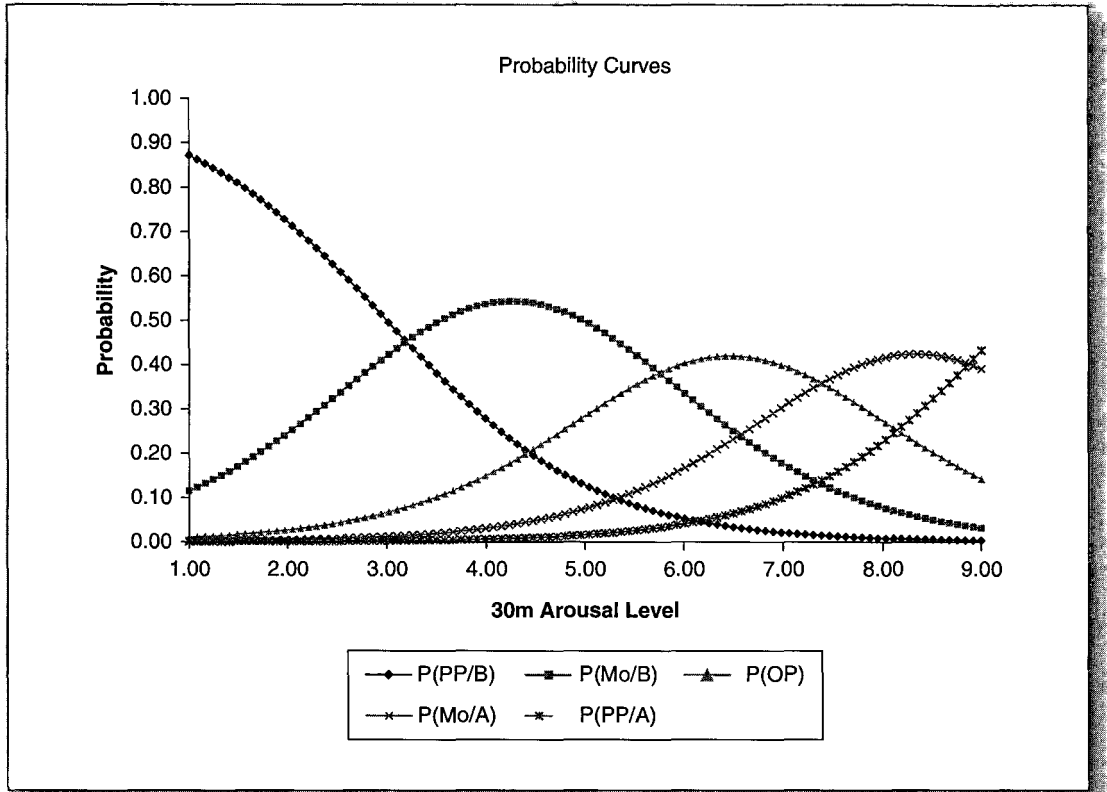


Figure X. IAPZ 5-category model for an archer derived from the ordinal logistic regression.

Figure D3 Sample Figure Representing Outcome Data

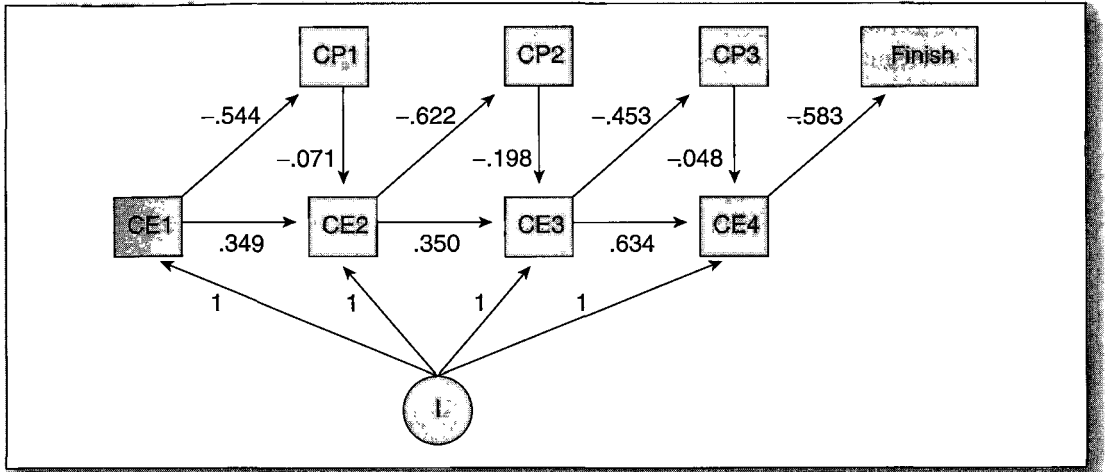


Figure X. Repeated measure structural model paths' coefficients. Correlation coefficients between collective efficacy (CE) and the subsequent performance times at each checkpoint (CP) revealed strong inverse relationships (e.g., $r_{CP1,CE1} = -.544$), although when factoring in CE with a previous measure of CE and a performance time the relationship weakened (e.g., $r_{CE2,CE1 \cdot CP1} = -.071$; $r_{CE2,CE1 \cdot T1} = .349$), though it was not relevant to the specific test of the stated hypotheses.

Figure D4 Sample Figure Representing Outcome Data

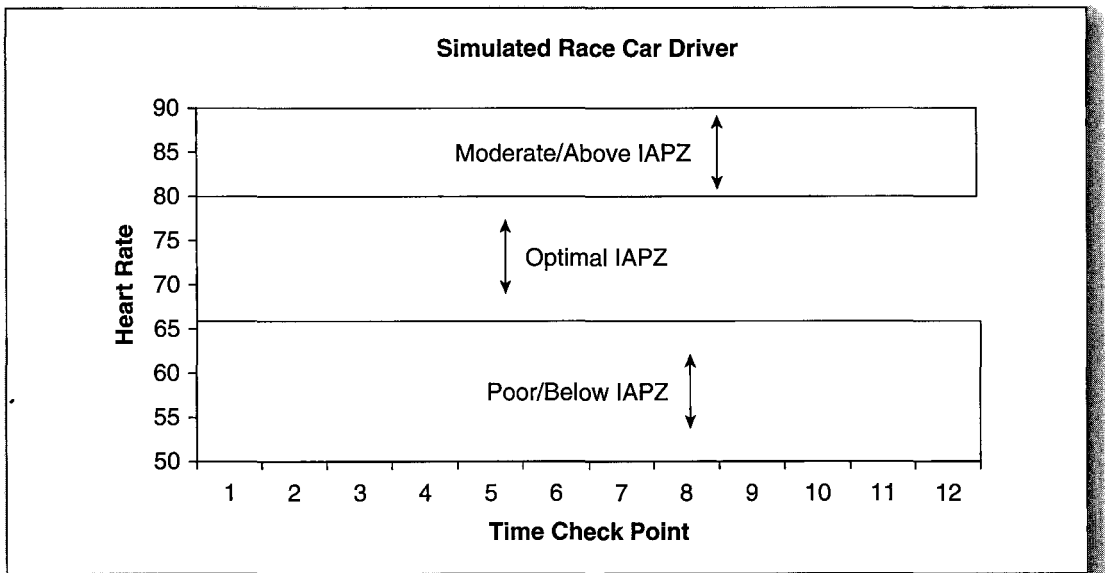


Figure X. Individual Affect-Related Performance Zone (IAPZ) profile charts for a simulated race car driver, which were derived from the IAPZ curves.

Summary Chart of Symbols

Symbol	Explanation
n	Sample size
N	Population or lot size
M	Mean
SD	Standard deviation of the sample
\bar{X}	Arithmetic mean of the sample
s	Standard deviation of the population
CV	Coefficient of variation
s^2	Sample variance
σ^2	Variance of the population
SE	Standard error
t	t distribution: shows both the random variable and a particular or observed value of this variable
F	F distribution: shows both the random variable and a particular or observed value of this variable
df	Degree(s) of freedom
χ^2	Chi-squared distribution: shows the random variable and a particular or observed value of this variable
P	Level of significance, probability
β	Regression coefficient of population
r	Coefficient of correlation, sample
r^2	Coefficient of determination for r
R	Coefficient of multiple correlation
R^2	Coefficient of determination for R
CI	Confidence interval
γ	Gamma
f	Frequency
τ_b	Kendall's tau- b
α	Level of significance

Symbol	Explanation
μ	Population mean score
σ^2	Population variance
ρ	Rho
z	z score
Z	Standardized scores, and Z distribution
r	Correlation
r_{pb}	Point biserial correlation
p	Statistical significance
Sig.	Statistical significance
95% CI	The 95% confidence interval
t	t test
F	One-way ANOVA or comparison test of between-group differences
eta	Correlation coefficient for ANOVA (3+ categories)
<i>Tukey-b</i>	ANOVA test of significance of difference among means
Beta	Standardized regression slope
R^2_{change}	Whether next variables entered add anything above variables entered
σ_x^2	Variance
σ_{xy}	Covariance (of x and y)

Appendix E

Analysis by Design

In this appendix, we present a statistics primer that will assist readers in reviewing the basic parametric statistical analysis associated with specific research designs presented in this book. This is not a complete guide to statistics. There are many terms and identifiers not included in this primer that should also be part of a particular analysis. We attempted to provide a quick example, including a vignette that details a research scenario so the reader can follow the steps of developing a research question (hypothesis) and then use the most appropriate research design and apply the necessary parametric statistical analysis to answer the stated questions. We also included a brief write-up of the results. However, we did not include data screening (e.g., tests for normality) examples prior to the analysis. In some instances, we provide data sets to accompany the analysis that can be accessed on the companion website. We also include a standardized chart that can be used to assess the extent that validity, control, and causal inferences were secured, considering all the circumstances. All data sets are completely fabricated and do not represent real data points.

We provide analysis examples for the following designs:

1 Pre- and Posttest Control Group Design

Grouping Variable (2 levels): One IV and one control group = Group 1 receives treatment, and Group 2 receives no treatment.

2 2-Factor Pre- and Posttest Design

Grouping Variable (2 levels): Two IVs and two groups = Group 1 receives Treatment A, and Group 2 receives Treatment B.

3 2-Factor Pre- and Double-Posttest Design

Grouping Variable (2 levels): One IV and one control group = Group 1 receives treatment, and Group 2 receives no treatment.

4 Crossover (3-Factor) Design

Grouping Variable (3 levels): Three IVs = Groups 1, 2, and 3 alternate between the three different treatments.

5 3 x 3 Latin-Square Design

Grouping Variable (3 levels): One IV = Group 1 receives Treatment Combo A, B, C; Group 2 receives Treatment Combo B, C, A; and Group 3 receives Treatment Combo C, A, B.

6 Wait-List Continuation Design (Switching-Replication Version With Random Assignment)

Grouping Variable (2 levels): One IV and one control group = Group 1 receives Treatment A; then conditions are switched, and Group 2 receives Treatment A.

7 Regression-Discontinuity Design

Grouping Variable (2 levels): One IV and one control group = Group 1 receives treatment, and Group 2 receives no treatment.

8 Ex Post Facto Design

Grouping Variable (2 levels): One IV and one control group = Group 1 and Group 2 receive the treatment.

Analysis and the Between-Subjects Approach

The pre- and posttest design is a common research design for the between-subjects approach. The pretest is collected for one of two reasons: (a) to measure the change from baseline (pretest to posttest differences or gain scores); or (b) to test for group equivalency. These designs are sometimes referred to as the *analysis of covariance designs*. Although these designs are considered common, ironically, there is little research and literature to support the most appropriate analytic strategies to handle data collected from this type of design. Many texts on statistics exemplify the application of an ANCOVA on observational or preexisting data sets, which is entirely nonexperimental. Furthermore, statistical software programs such as SPSS do not have an obvious designation to handle a pre- and posttest design. Therefore, separate data analyses must be performed and combined to address the research questions of interest.

The primary statistics for all the pre- and posttest designs listed is the ANCOVA, with slight variations based on the grouping variable (the grouping variable is the categorical variable otherwise known as a *fixed or random factor* in statistical programs). Why ANCOVA? The pretest measure should be correlated with the posttest measure (because it is the same measure taken at least twice in sequence). Thus, the posttest measure will be assigned as the DV, the grouping variable will be assigned as the IV (although it is technically *not* the IV), and the pretest becomes the covariate.

The second approach to analyzing data from pre- and posttest designs is to test the differences in the “gain” scores, which is basically the score obtained from subtracting the pretest results from the posttest. This can be achieved through the use of an independent *t* test (two groups) or a one-way ANOVA when three or more groups are involved. The gain score analysis (GSA) should be conducted when the researcher wants to fully understand the effect of the treatment from pre- to posttest. The ANCOVA should be used when the researcher is just interested in determining the effect of the treatment on the posttest, which otherwise could not be predicted via the pretest measures (see Knapp & Schafer, 2009, for more on GSA). The two-way mixed ANOVA can also be used for these designs but is not recommended due to various limitations in the analysis.

ASSUMPTIONS

1. Normality—If data are normal, then proceed to the parametric analysis. If data are not normal, then proceed to a nonparametric procedure.
2. Homogeneity of Variance—See options within each analysis to determine homogeneity of variance.
3. Independence—It is during the data collection procedures that independence is secured.

ANCOVA-SPECIFIC ASSUMPTIONS

1. A linear relationship exists between the posttest and the pretest (i.e., homogeneity of regression assumption). If violated, then refer to the Johnson-Neyman technique.
2. The covariate is measured without error.

Analysis and the Within-Subjects Approach

The within-subjects approach is another common approach used in education and the social, behavioral, and health sciences. The general premise is that one group (or subject) serves in each of the treatment conditions. The most critical aspect to consider with this approach is that performance in one treatment condition affects the performance in a second treatment condition; therefore, it is vital to randomize the order of the treatments (also known as *counterbalancing*) to control for sequencing effects (a major threat to internal validity).

Explicitly stated throughout this book, the research question drives the type of design and then the appropriate analysis ensues. Most of the research designs for the within-subjects approach require the repeated-measures ANOVA (RM ANOVA). If the general assumptions are not met (e.g., normality and sphericity), which they rarely are, then a strong alternative analysis is the MANOVA (see O'Brien & Kaiser, 1995). If multiple dependent variables are included, then the multivariate versions should be used anyway (MANOVA), and in some cases, the hierarchical linear model (HLM) can be applied. Data screening is paramount for this approach to determine the level of association that exists between the dependent variables.

Why repeated-measures analysis? If a research question calls for the examination and collection of data points over time, then the RM ANOVA is applied to either (a) assess changes in statistical values over three or more time points or (b) assess the differences in statistical values within three or more conditions. The RM ANOVA is an extension of the dependent-samples *t* test, sometimes referred to as the *within-subject ANOVA*.

REPEATED-MEASURES ASSUMPTIONS

1. **Multivariate Normality**—If data are normal, then proceed to the parametric analysis. If data are not normal, then proceed to the nonparametric procedure.
2. **Sphericity or Circularity**—See options within each analysis to determine sphericity and compound symmetry. If sphericity is violated, then alternative analysis should be used, such as MANOVA or an adjusted univariate test (see Randall, R. R., & Barcikowski, 1987).
3. **Independence**—It is during the data collection procedures that independence is secured.

Example E1 *Pre- and Posttest Control Group Design*

A researcher is interested in examining the impact of a behavioral-therapy intervention on levels of depression in patients who are clinically depressed. The behavioral-therapy intervention to be used is an established approach and requires the patient to participate in a total of nine treatment sessions over a 3-week period. In order to assess the impact of the intervention, one group will receive the therapy, and the second group will receive therapeutic attention and will be considered an attention-control group. They will be assessed before and immediately following the intervention. A nonprobability purposive sampling technique will be used to identify 100 patients diagnosed with clinical depression and then will be randomly assigned to either the treatment or control group.

General Research Question: To what extent do those who receive the treatment (Group 1) differ from those who do not receive treatment (Group 2) on their outcome (DV: continuous data)?

Grouping Variable (2 levels): One IV and one control group = Group 1 receives treatment (behavioral therapy), and Group 2 receives no treatment (attention-control group).

DV: Continuous (BDI)

Group	Pretest	Treatment	Posttest
1	O ₁	X	O ₂
2	O ₁	–	O ₂

Time ►

Note: Random assignment to conditions must be used to qualify as experimental research. If no random assignment was used, then this is not the appropriate analysis (see reliability-corrected ANCOVAs for nonequivalent group designs or quasi-experimental research).

Design: Experimental research using a between-subjects approach with a pre- and posttest control group design

Null Hypothesis: The adjusted population means for all groups are equal. In this example, $H_0: \mu_1 = \mu_2$ or the adjusted means of Groups 1 and 2 are equivalent.

Statistic: Analysis of Covariance (ANCOVA)

Statistical Power: Utilize G*Power Data Analysis

Note: The ANCOVA adjusts the posttest *means* to what they would be if all groups were equivalent on the covariate (the pretest) and at the grand *mean* (the *mean* of the *means*).

Specific Research Question: To what extent do those who receive behavioral therapy differ from those who do not receive behavioral therapy on their depression scores?

Null Hypothesis: There is no difference in the scores on the Beck's Depression Inventory (BDI) for those who did receive behavioral therapy compared to those who did *not* receive behavioral therapy.

<i>Group</i>	<i>Assignment</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
1 (<i>n</i> = 50)	R	BDI	Behavioral Therapy	BDI
2 (<i>n</i> = 50)	R	BDI	–	BDI

Additional Notes: In this particular example, we will assume the control group is an attention control group, and the behavioral therapy intervention is delivered via established approaches with the use of a manipulation check to enhance design controls (i.e., manipulation and elimination).

DATA ENTRY

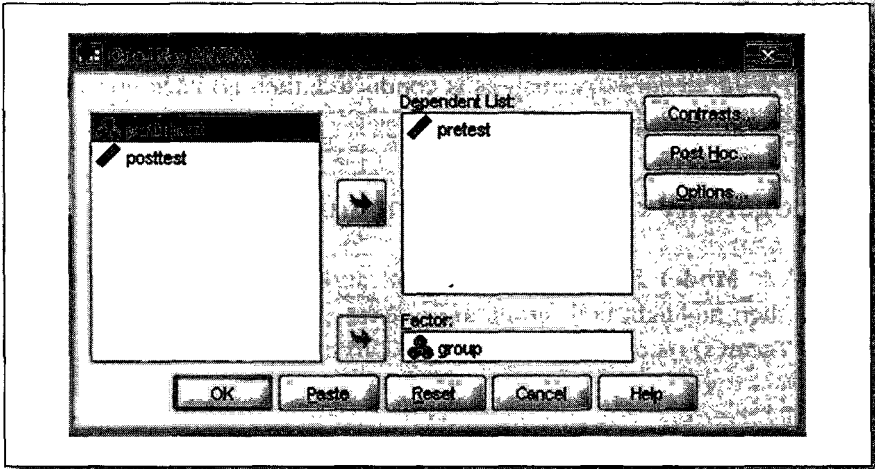
See data set for this particular example for an understanding on how to set up Variables in Variable view and how to enter data into SPSS for the between-subjects approach and pre- and posttest designs.

♦ COMMAND PROMPTS FOR SPSS

Data titled *PrePostControl.sav*

Homogeneity of Variances and Descriptives for Pretest Data

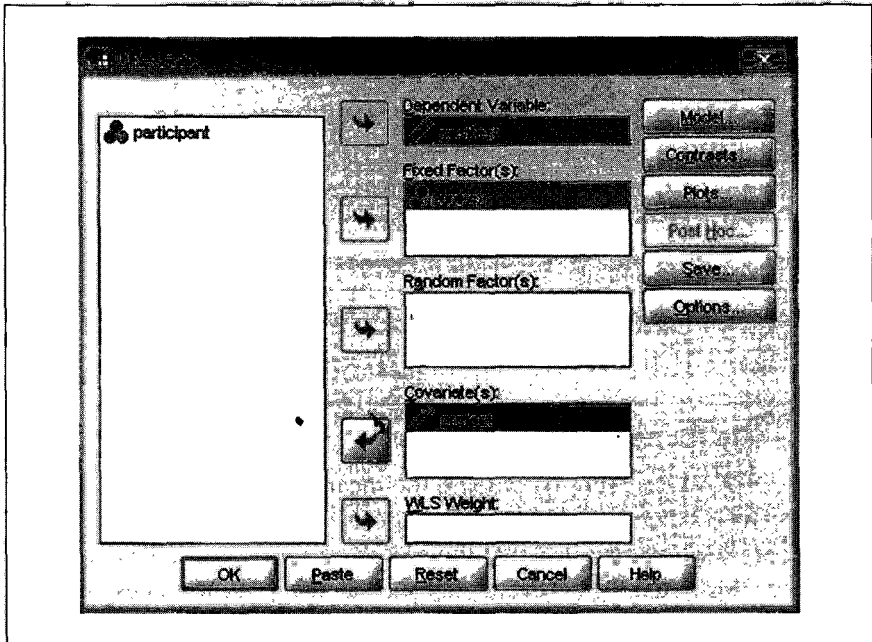
Analyze → Compare Means → One-Way ANOVA



Select **Options**, then select *Homogeneity of variance tests*, and *Descriptive*.

Ancova Including Effect Size and Parameter Estimates

Analyze → General Linear Model → Univariate

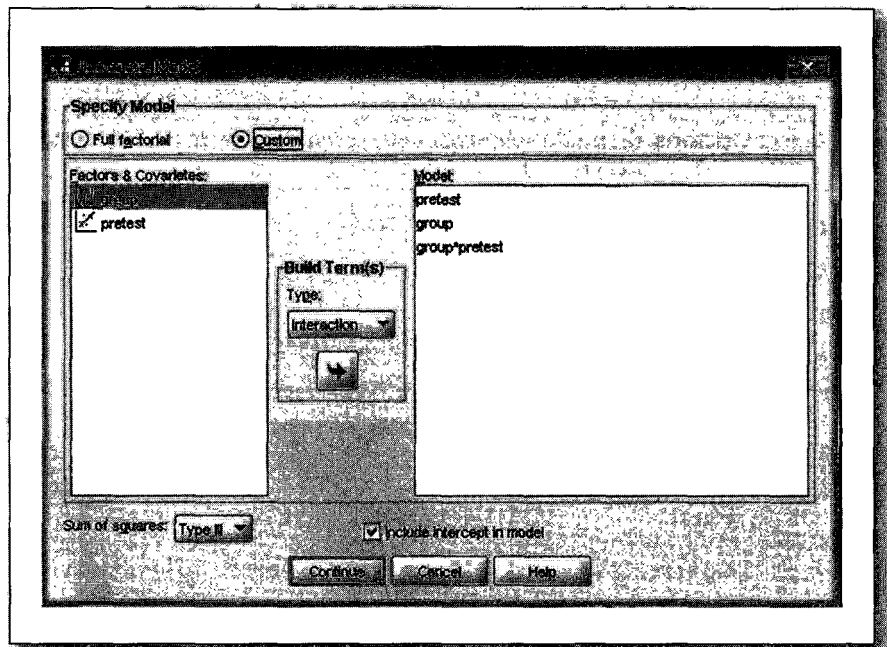


Select **Options** then select *Estimates of effect size*, *Observed power*, *Parameter estimates*, and *Descriptive*.

Once the preliminary analysis is conducted, then go back and include the following:

Homogeneity of the Coefficients for the Pretest

Select **Model** from **Univariate** box and carryover *pretest* and the *group*, then highlight both *group* and *pretest*, then click **Interaction** from **Build Term(s)** menu and click arrow to carryover to **Model**.



Note: The main analysis is conducted *without* the Custom option selected in the Specify Model box.

Results

An ANCOVA was run to compare the mean BDI scores for participants who received behavioral therapy and those who received no treatment. Preliminary analysis revealed that the scores on the pretest measures were normally distributed. The Levene statistic on the pretest scores revealed nonsignificance $F(1, 98) = .221, p > .639$, indicating homogeneity of variances (i.e., equal variances between the groups). Further, the interaction for the group assignment and pretest measures was not significant ($p = .791$),

Test of Homogeneity of Variances

pretest			
Levene Statistic	df1	df2	Sig.
.221	1	98	.639

The Levene statistic is not significant, indicating homogeneity of variances.

Descriptives

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
treatment group	50	31.7000	6.12208	.86579	29.9801	33.4399	18.00	45.00
control group	50	33.7200	5.92828	.83894	32.0362	35.4048	22.00	48.00
Total	100	32.7100	6.06076	.60808	31.5034	33.9188	18.00	48.00

Means and standard deviations illustrating the differences pre- to posttest

Descriptive Statistics

Dependent Variable: posttest			
group	Mean	Std. Deviation	N
treatment group	23.5000	6.38637	50
control group	29.0200	8.63900	50
Total	26.2600	8.05112	100

Partial Eta Squared Effect Size
 $\eta^2 = .01$ — "small" effect
 $\eta^2 = .09$ — "medium" effect
 $\eta^2 = .25$ — "large" effect

group * pretest	4.136	1	4.136	.071	.791	.001	.071	.058
Error	5929.120	98	58.637					
Total	75376.000	100						
Corrected Total	6417.240	99						

Note: These are the results from running the Custom option in the Modeling function. Discard the remainder of the results, and run the main analysis again by selecting the Full factorial function.

The p value shows nonsignificance, indicating homogeneity of regression

Tests of Between-Subjects Effects

Dependent Variable: posttest								
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power ^a
Corrected Model	763.984 ^a	2	381.992	6.750	.002	.122	13.500	.910
Intercept	2679.215	1	2679.215	48.117	.000	.322	48.117	1.000
pretest	22.224	1	22.224	.383	.539	.004	.383	.094
group	763.982	1	763.982	13.500	.000	.122	13.500	.893
Error	5693.256	97	58.075					
Total	75376.000	100						
Corrected Total	6417.240	99						

a. R Squared = .122 (Adjusted R Squared = .104)
 b. Computed using alpha = .05

Differences between groups are shown to be significant with a medium effect size.

Parameter Estimates

Dependent Variable: posttest							
Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Partial Eta Squared
					Lower Bound	Upper Bound	
Intercept	31.685	4.440	7.135	.000	22.872	40.498	.344
pretest	-.079	.128	-.619	.539	-.333	.175	.004
[group=1.00]	-5.680	1.546	-3.674	.000	-8.748	-2.612	.122
[group=2.00]	0 ^a						

a. This parameter is set to zero because it's redundant.

Those who received behavioral therapy over no therapy reduced their scores on the BDI by 5.68 points.

with a negligible effect size ($\eta^2 = .001$), which also indicates no substantial group differences on the pretest measures (i.e., the homogeneity of regression assumption is not violated). The differences between the therapy and control group were revealed to be statistically significant, $F(1, 97) = 13.50$, $p < .05$, partial $\eta^2 = .12$. The BDI scores on the pretest for the control group were minimally higher ($M = 33.72$, $SD = 5.928$) than the posttest scores ($M = 29.02$, $SD = 8.64$), indicating a minor change. For the treatment group, the BDI scores on the pretest were higher ($M = 31.70$, $SD = 6.12$) than the posttest scores ($M = 23.50$, $SD = 6.38$), indicating a “medium” effect as revealed by the effect size ($\eta^2 = .12$). Specifically, the parameter estimates indicate that on average individuals who receive the behavioral therapy would reduce their depression scores on the BDI by 5.68 points.

<i>Pre- and Posttest Control Group Design: Validity, Control, and Causal Inferences</i>		
	<i>Level</i>	<i>Explanation</i>
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
History	↑	Inclusion of pretest
Maturation	↑	Inclusion of pretest
Testing	↑	Inclusion of pretest
Instrumentation	↓	The instrument did not change between pre- and posttest
Statistical regression	↔	Random assignment applied
Attrition	↓	No participants dropped out following the pretest
Selection bias	↓	Random assignment applied
Combination of selection and other treatments	↓	Random assignment and independent groups
Diffusion	↓	Conditions are independent
Special treatment	↓	No-contact control group
Sequencing effects	n/a	Between-subjects approach
Statistical Conclusion Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Low statistical power	↓	Adequate sample size
Assumption violation of statistical tests	↓	Normality, homogeneity of variance, independence, linear relationship between post- and pretest confirmed
Error rate problem	↓	Single primary analysis used

Restriction of range	↓	Treatment and control conditions
Extraneous variance in the experimental setting	↔	Strong levels of design control implemented
Inaccurate effect size estimation	↓	One-way ANCOVA is the most appropriate analysis
Variability in procedures	↓	Strong levels of design control implemented
Subject heterogeneity	↓	Random selection and random assignment
Unreliability of the measures	↓	BDI maintains adequate psychometric properties
Multiple comparisons and error rates	↓	Single dependent variable and one primary analysis used
Control—Design and statistical control is (strong †††) (medium ††) (weak †)		
Manipulation	†††	Established therapeutic approach w/ manipulation check
Elimination	†††	Established therapeutic approach w/ manipulation check
Inclusion	†††	Multiple groups compared (included treatment and control)
Group or condition assignment	†††	Random assignment
Statistical procedures	††	Pretest entered as covariates
Cause and Effect—Cause-effect determination is (strong †††) (medium ††) (weak †)		
Covariation	†††	Strong controls and minimal threats to internal and statistical conclusion validity
Temporal precedence	†††	Strong controls and minimal threats to internal and statistical conclusion validity
No plausible alternative explanations	††	Covariation and temporal precedence soundly confirmed, but replicable results should be obtained over time

Note: All threats are not created equal. Threats to construct and external validity should also be considered.

Example E2 2-Factor Pre- and Posttest Design

A researcher is interested in determining the effectiveness of two different interventions on emotional intelligence for children in middle school. The first intervention is an established in-class social-skills training and includes a protocol for implementation. The second intervention is a parent-centered intervention and requires the parents to be trained on how to nurture and improve emotional intelligence in their children. Each program

is design to cover a 4-week span. A nonprobability purposive sampling technique will be used to identify 80 middle school children. They will then be randomly assigned to one of two treatment conditions. Both interventions will follow an established standardized protocol for implementation and will use manipulation checks to ensure the programs are implemented as intended to ensure the integrity of various design controls (e.g., manipulation and elimination).

General Research Question: To what extent do those who receive Treatment A (Group 1) differ from those who receive Treatment B (Group 2) on their outcome (DV: continuous data)?

Grouping Variable (2 levels): Two IVs = Group 1 receives Treatment A (in-class social skills trainings) and Group 2 receives Treatment B (parent-centered EQ training)

DV: Continuous (EQ-i)

<i>Group</i>	<i>Pretest</i>	<i>Treatment</i>	<i>Posttest</i>
1	O ₁	X _A	O ₂
2	O ₁	X _B	O ₂

Time ►

Note: Random assignment to conditions must be used to qualify as experimental research. If no random assignment was used, then this is not the appropriate analysis (see reliability-corrected ANCOVAs for nonequivalent group designs or quasi-experimental research).

Design: Experimental research using a between-subjects approach with a 2-factor pre- and posttest design

Null Hypothesis: The adjusted population means for all groups are equal. In this example, $H_0: \mu_1 = \mu_2$ or the adjusted means of Groups 1 and 2 are equivalent.

Statistic: Gain Score Analysis (GSA; independent t test)

Statistical Power: Utilize G*Power Data Analysis

Specific Research Question: To what extent do those who receive the in-class social-skills training differ from those who receive parent-centered EQ training on their emotional intelligence?

Null Hypothesis: There is no difference in the scores on the Emotional Intelligence test (BarOn EQ-i) for those who receive the in-class social-skills training compared to those who receive the parent-centered EQ training.

Group	Assignment	Pretest	Treatment	Posttest
1 (n = 40)	R	EQ-i	In-class social skills training	EQ-i
2 (n = 40)	R	EQ-i	Parent-centered EQ training	EQ-i

A gain score analysis (GSA) is used for the following example. A GSA should only be used if random assignment is utilized, thus assuming each group (condition) is considered probabilistically “equivalent.” However, based on theoretical considerations, an ANCOVA can also be applied. See the previous example for details on how to apply an ANCOVA, as the procedures are the same.

DATA ENTRY

See data set for this particular example for an understanding on how to set up variables in Variable View and how to enter data into SPSS for the between-subjects approach and pre- and posttest designs.

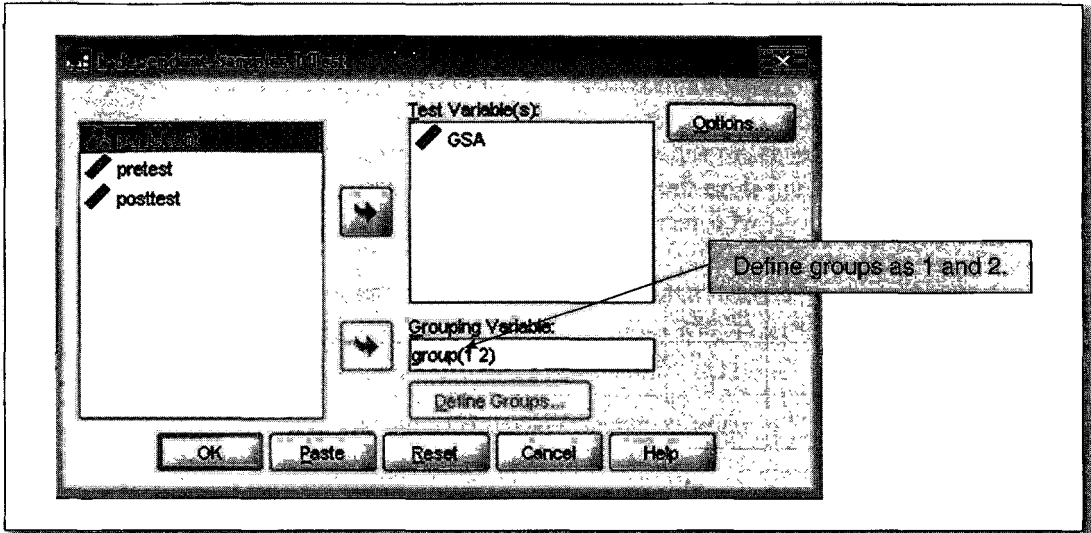
COMMAND PROMPTS FOR SPSS ♦

Data titled *2-factorPrePosttest.sav*

Independent samples *t* test

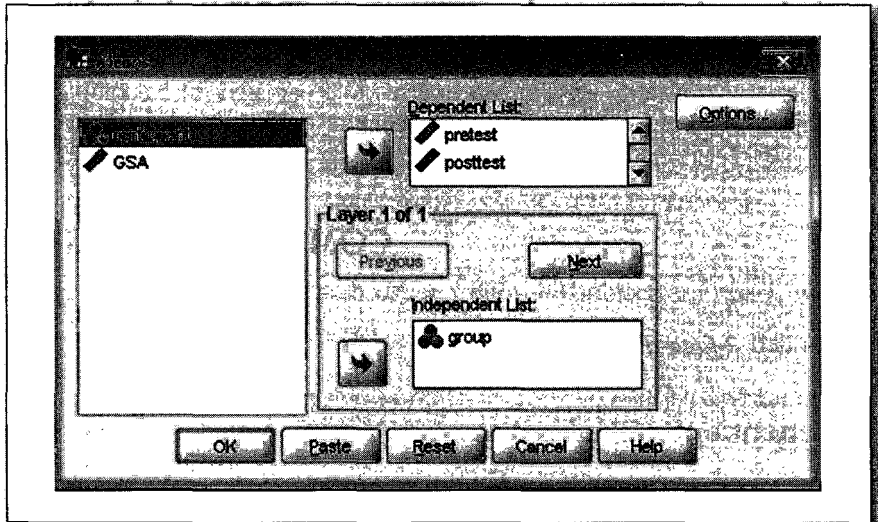
Analyze → Compare Means → Independent Samples *t* Test

Note: Create a column in Excel titled *GSA*, which are the gain scores (posttest minus the pretest).



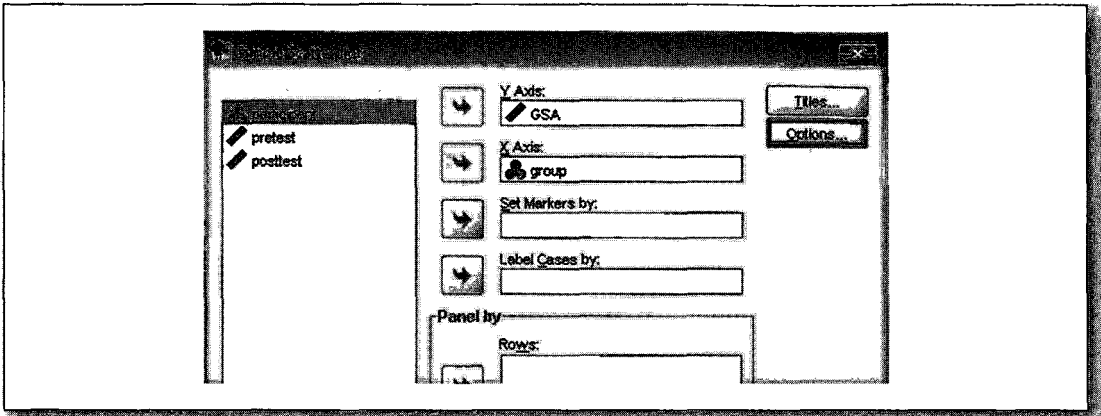
Descriptive Statistics for the Pre- and Posttest Scores

Analyze → Compare Means → Means



Point Biserial Correlation Coefficient (Effect Size)

Graph → Legacy Dialogues → Scatter/Dot → Simple Scatter



Note: To run the r_{pb} , select **Analyze** → **Correlate** → **Bivariate**, and in the Variables box enter *GSA* and *group*.

The Levene statistic is not significant, indicating homogeneity of variances.

Mean gain scores between the two treatment groups

Treatment A on average scored 5.72 higher than Treatment B.

The *t* statistic shows significance ($p < .05$).

Mean scores on the EQ-1 between the pre- and posttests for both treatment groups

Group Statistics

group	N	Mean	Std. Deviation	Std. Error Mean
GSA TreatmentA	40	14.3000	8.66145	1.52761
TreatmentB	40	8.5750	11.22928	1.77551

Independent Samples Test

GSA	Equal variances assumed	Levene's Test for Equality of Variances		t-test for Equality of Means		95% Confidence Interval of the Difference			
		F	Sig.	t	Sig. (2-tailed)	Lower	Upper		
EQ-1	Equal variances assumed	1.593	.208	2.444	.017	2.72590	2.34222	1.00199	19.20091
	Equal variances not assumed			2.444	.017	2.72590	2.34222	1.06235	19.20095

Report

group		pretest	posttest
TreatmentA	Mean	96.5000	110.8000
	N	40	40
	Std. Deviation	6.19348	6.92524
TreatmentB	Mean	85.6250	104.2000
	N	40	40
	Std. Deviation	5.26093	10.28367
Total	Mean	96.0625	107.5000
	N	80	80
	Std. Deviation	5.72656	8.32283

Pooled and Separate Variances
If the Levene test is nonsignificant, then the "Equal variances assumed" row, which are the results from the pooled-variances *t* test, should be used. If the results are significant, then the results from the "Equal variances not assumed" row should be used. These are the results from the separate-variances *t* test, which is known as the Welch-Satterthwaite *t* test (see Zimmerman and Zumbo, 2009, for considerations with these tests).

Effect Size

SPSS does not provide a specific effect size calculation for the independent samples t test. The appropriate formula for the independent samples t test is the Cohen's d , which uses the pooled standard deviation and is also known as Hedges' g .

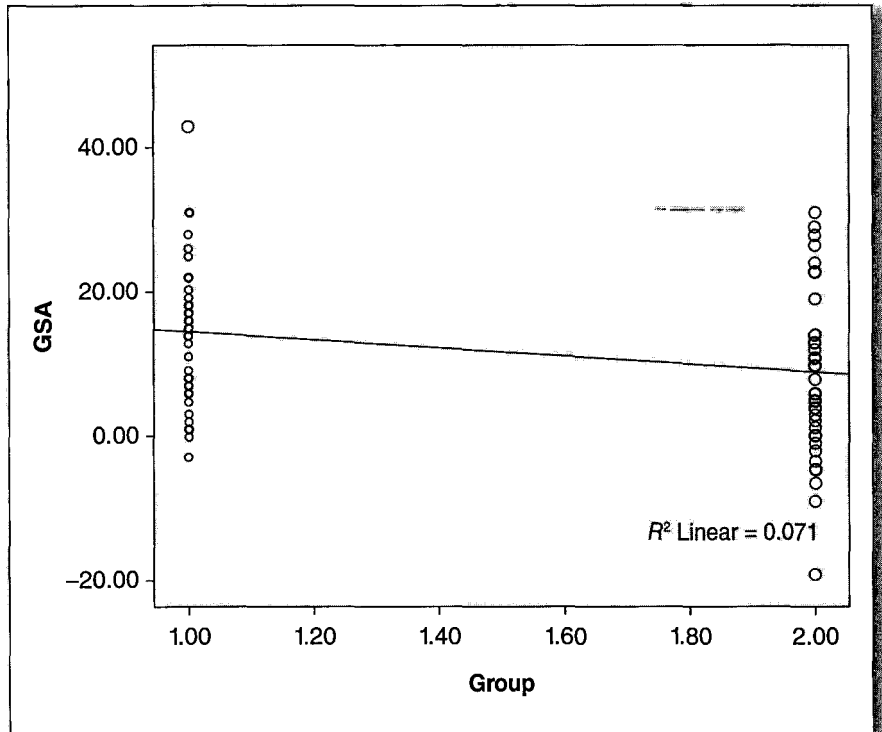
$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

$$d = \frac{14.3 - 8.575}{\sqrt{\frac{(39)(9.661)^2 + (39)(11.229)^2}{40 + 40 - 2}}} = \frac{5.725}{10.474} = 0.5466 = .55$$

Cohen's d

$d = .20$ —"small" effect
 $d = .50$ —"medium" effect
 $d = .80$ —"large" effect

The second option for an effect size calculation for the independent samples t test is the point biserial correlation.



Note: Double click on the chart in SPSS output, and then click on "Add Fit Line at Total" to get the regression line and R^2 value.

Results

An independent samples *t* test was used to compare the gain scores on the BarOn EQ-i test for participants who received in-class social skills training and those who received parent-centered EQ training. The difference between the two conditions were found to be statistically significant $t(78) = 2.44, p < .017$. These findings indicate that individuals who received in-class social skills training (Treatment A) on average scored higher ($M = 14.3, SD = 9.66$) than those who received the parent-centered EQ (Treatment B) ($M = 8.57, SD = 11.22$). The *Mean* difference between the two groups was 5.73. The effect size calculation revealed a “medium” effect ($d = .55$). In addition, the point biserial correlation revealed a “medium” effect ($r_{pb} = -.27; R^2 = -.071$).

2-Factor Pre- and Posttest Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
History	↑	Inclusion of pretest
Maturation	↑	Inclusion of pretest
Testing	↑	Inclusion of pretest
Instrumentation	↓	The instrument did not change between pre- and posttest
Statistical regression	↔	Random assignment applied
Attrition	↓	No participants dropped out following the pretest
Selection bias	↓	Random assignment applied
Combination of selection and other treatments	↓	Random assignment and independent groups
Diffusion	↓	Conditions are independent
Special treatment	↓	Independent treatment conditions
Sequencing effects	n/a	Between-subjects approach
Statistical Conclusion Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Low statistical power	↘	Adequate sample size
Assumption violation of statistical tests	↓	Normality, homogeneity of variance, and independence confirmed
Error rate problem	↓	Single primary analysis used
Restriction of range	↓	Distinct treatment conditions

(Continued)

(Continued)

Extraneous variance in the experimental setting	↔	Strong levels of design control implemented
Inaccurate effect size estimation	↔	GSA does not adjust for pretest differences
Variability in procedures	↓	Strong levels of design control implemented
Subject heterogeneity	↓	Random selection and random assignment
Unreliability of the measures	↓	EQ-i maintains adequate psychometric properties
Multiple comparisons and error rates	↓	Single dependent variable and one primary analysis used
Control —Design and statistical control is (strong TTT) (medium TT) (weak T)		
Manipulation	TTT	Established therapeutic approaches w/ manipulation check
Elimination	TTT	Established therapeutic approaches w/ manipulation check
Inclusion	TTT	Multiple groups compared (two treatment conditions)
Group or condition assignment	TTT	Random assignment
Statistical procedures	TT	Gain scores established (posttest minus pretest)
Cause and Effect —Cause-effect determination is (strong TTT) (medium TT) (weak T)		
Covariation	TTT	Strong controls and minimal threats to internal and statistical conclusion validity
Temporal precedence	TTT	Strong controls and minimal threats to internal and statistical conclusion validity
No plausible alternative explanations	TT	Covariation and temporal precedence soundly confirmed, but replicable results should be obtained over time

Note: All threats are not created equal. Threats to construct and external validity should also be considered.

Example E3 2-Factor Pre- and Double-Posttest Design

A researcher is interested in comparing the residual or lasting effects of two different types of therapy on perceived ability to “function” in patients who have recently undergone chemotherapy. Treatment A is a biofeedback (BFB) therapy and will be implemented over the course of 6 weeks. Treatment B is a yoga program specifically designed for cancer patients and will also be implemented over a 6-week span. Through the use of a

nonprobability purposive sampling technique, a total of 50 participants have been identified who have recently received their first rounds of chemotherapy and will be randomly assigned to one of the two conditions. Both interventions will follow an established standardized protocol for implementation and will use manipulation checks to ensure the programs are implemented as intended to ensure the integrity of various design controls (e.g., manipulation and elimination).

General Research Question: To what extent do those who receive Treatment A (Group 1) differ from those who receive Treatment B (Group 2) on their outcome (DV: continuous data) measured over time?

Grouping Variable (2 levels): One IV and one control group = Group 1 receives Treatment A (biofeedback therapy) and Group 2 receives Treatment B (yoga)

DV: Continuous (QOLFS)

Group	Pretest	Treatment	Posttest1	Posttest2
1	O ₁	X _A	O ₂	O ₃
2	O ₁	X _B	O ₂	O ₃

Time ►

Note: Although this design is listed under the within-subjects approach, the between-subjects differences are just as relevant to the study as the within-subject variances. Any number of posttests and factors can be included, based on theoretical and logistical considerations.

Design: Experimental research using a mixed-subjects (between and within) approach with a 2-factor pre- and double posttest design

Null Hypothesis: The adjusted population means for both groups are equal. In this example, $H_0: \mu_1 \approx \mu_2$ and $\mu_3 = \mu_4$ or the adjusted means of Groups 1 and 2 are equivalent.

Statistic: Repeated Measures Analysis of Variance (RM-ANOVA)

Statistical Power: Utilize G*Power Data Analysis

Specific Research Question: To what extent do those who receive biofeedback therapy differ from those who attend yoga therapy on their levels of perceived functioning over time?

Null Hypothesis: Over time, there is no difference in the scores on the Quality of Life Functioning Scale (QOLFS) for those who received biofeedback therapy compared to those who participated in the yoga program.

Group	Assignment	Pretest	Treatment	Posttest1	Posttest2
1 (n = 25)	R	QOLFS	BFB therapy	QOLFS	QOLFS
2 (n = 25)	R	QOLFS	Yoga	QOLFS	QOLFS

Note: The general premise for including a follow-up posttest is to examine the lasting effects of the therapies. The decision to consider at what time points the posttests are taken should be considered, based on theoretical and logistical tenants.

DATA ENTRY

See data set for this particular example for an understanding on how to set up Variables in Variable view and how to enter data into SPSS for the within-subjects approach.

♦ COMMAND PROMPTS FOR SPSS

Data titled *Predoublepost.sav*

RM ANOVA

Analyze → General Linear Models → Repeated Measures

Repeated Measures Define Factor(s)

Within-Subject Factor Name:

Number of Levels:

Add
Change
Remove

Time_Points(3)

Measure Name:

Add
Change
Remove

QOLFS

Define **Reset** **Cancel** **Help**

Enter the name of the WS factor as Time_Points, and enter the number of time points, then select Add; enter the name of the DV in the Measure Name, and select Add; select Define.

Repeated Measures

participant

Within-Subjects Variables
(Time_Points):
pre(1,QOLFS)
post1(2,QOLFS)
post2(3,QOLFS)

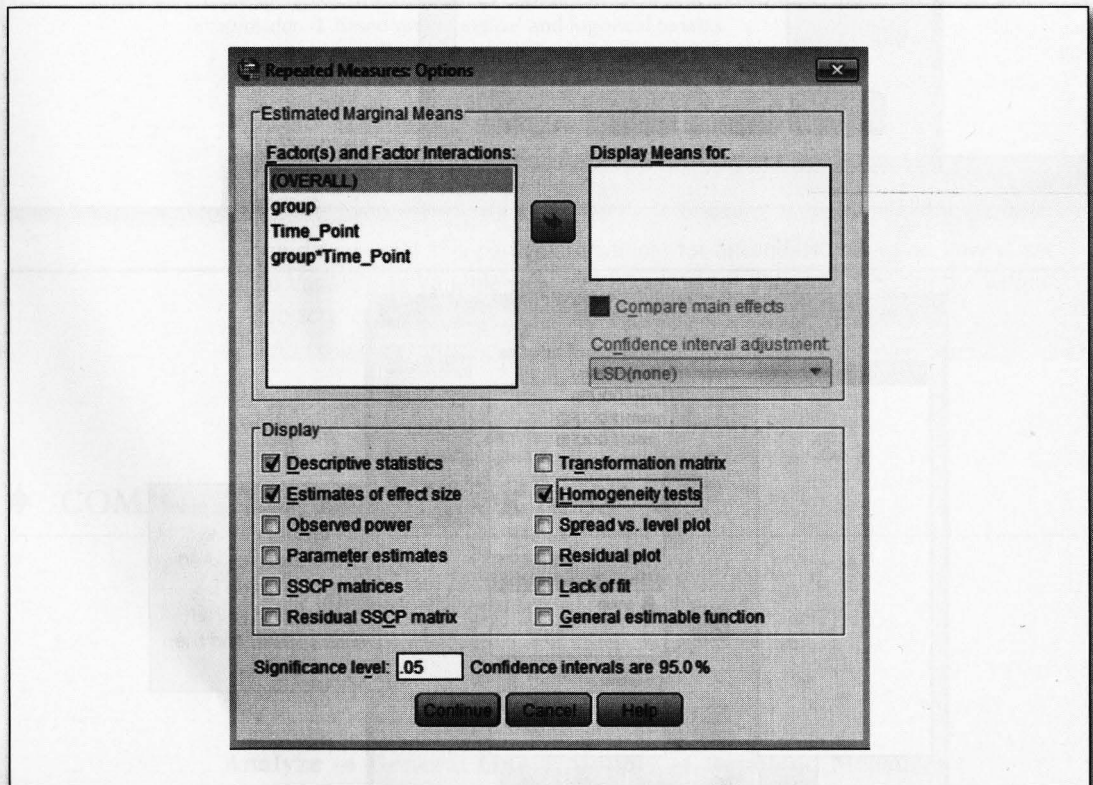
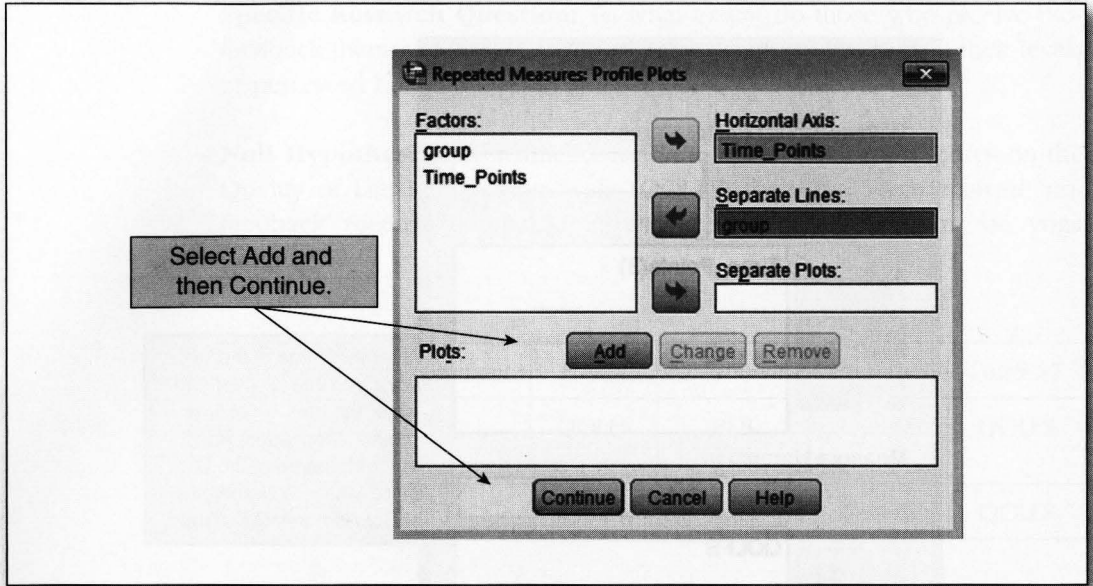
Between-Subjects Factor(s):
group

Covariates:

Model
Contrasts
Plots
Post Hoc
Save
Options

OK **Paste** **Reset** **Cancel** **Help**

Enter each time point in order in the WS Variables box. Add *group* to the BS Factors box, then select Plots, and then Options.



Descriptive Statistics

	group	Mean	Std. Deviation	N
pre	BFB Therapy	3.8800	1.26886	25
	Yoga	4.0400	.97809	25
	Total	3.9600	1.12413	50
post1	BFB Therapy	6.8800	.97125	25
	Yoga	4.9200	1.11505	25
	Total	5.9000	1.43214	50
post2	BFB Therapy	7.1200	.97125	25
	Yoga	5.6000	1.08012	25
	Total	6.3600	1.27391	50

Box's Test of Equality of Covariance Matrices^a

Box's M	3.401
F	.528
df1	6
df2	16693.132
Sig.	.787

Box's M test is not significant, indicating multivariate normality.

Box's M tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

Mauchly's Test of Sphericity^a

Measure: QOLFS	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Time_Point	.977	1.077	6	.986	.978	1.000	.500

Mauchly's test is not significant, indicating sphericity.

Tests the null hypothesis that the error covariance matrix of the standardized transformed dependent variables is proportional to an identity matrix.

Mauchly's test is not significant; therefore, the "Sphericity Assumed" row should be used.

Tests of Within-Subjects Effects

Measure: QOLFS	Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time_Point	Sphericity Assumed	162.253	2	81.127	64.301	.000	.573
	Greenhouse-Geisser	162.253	1.958	82.984	64.301	.000	.574
	Huynh-Feldt	162.253	2.000	81.127	64.301	.000	.573
	Lower-bound	162.253	1.000	162.253	64.301	.000	.573
Time_Point * group	Sphericity Assumed	31.293	2	15.647	12.402	.000	.205
	Greenhouse-Geisser	31.293	1.958	15.801	12.402	.000	.205
	Huynh-Feldt	31.293	2.000	15.647	12.402	.000	.206
	Lower-bound	31.293	1.000	31.293	12.402	.001	.205
Error(Time_Point)	Sphericity Assumed	121.120	96	1.262			
	Greenhouse-Geisser	121.120	93.974	1.290			
	Huynh-Feldt	121.120	96.000	1.262			
	Lower-bound	121.120	48.000	2.523			

Tests of Within-Subjects Contrasts

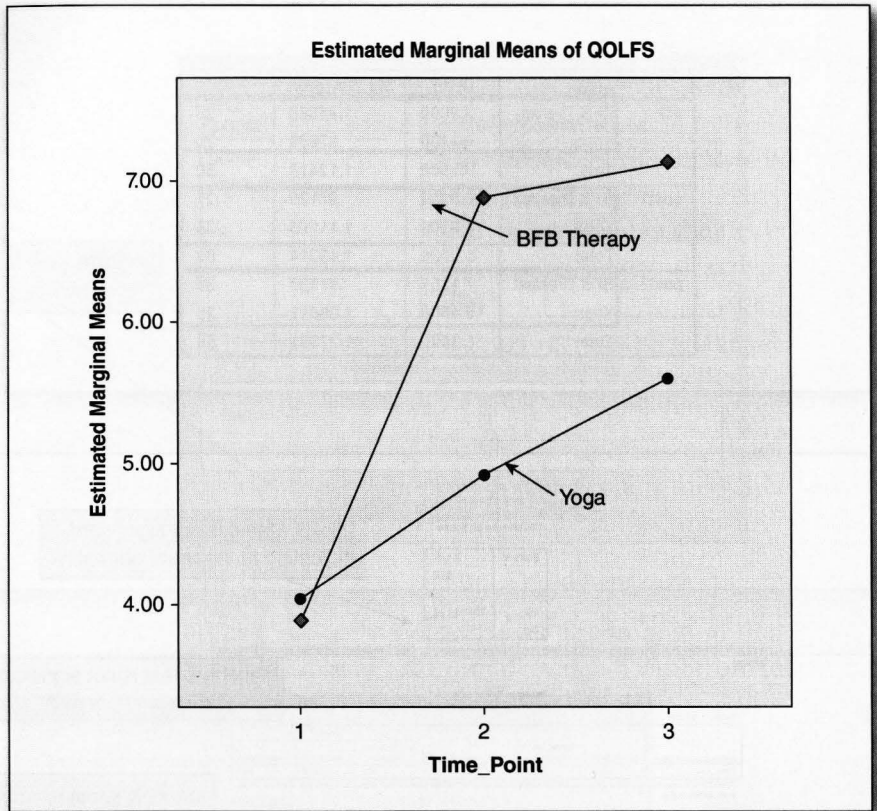
Measure: QOLFS	Source	Time Point	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time_Point	Linear	144.000	1	144.000	116.437	.000	.712	
		18.253	1	18.253	13.960	.000	.225	
Time_Point * group	Linear	17.640	1	17.640	14.509	.000	.232	
		13.653	1	13.653	10.442	.002	.179	
Error(Time_Point)	Linear	58.360	48	1.216				
		62.760	48	1.308				

Tests of Between-Subjects Effects

Measure: QOLFS

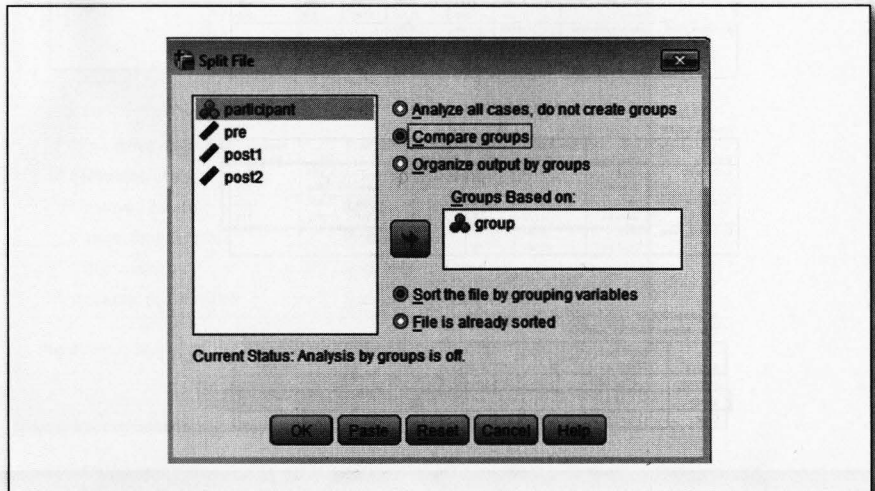
Transformed Variable: Average

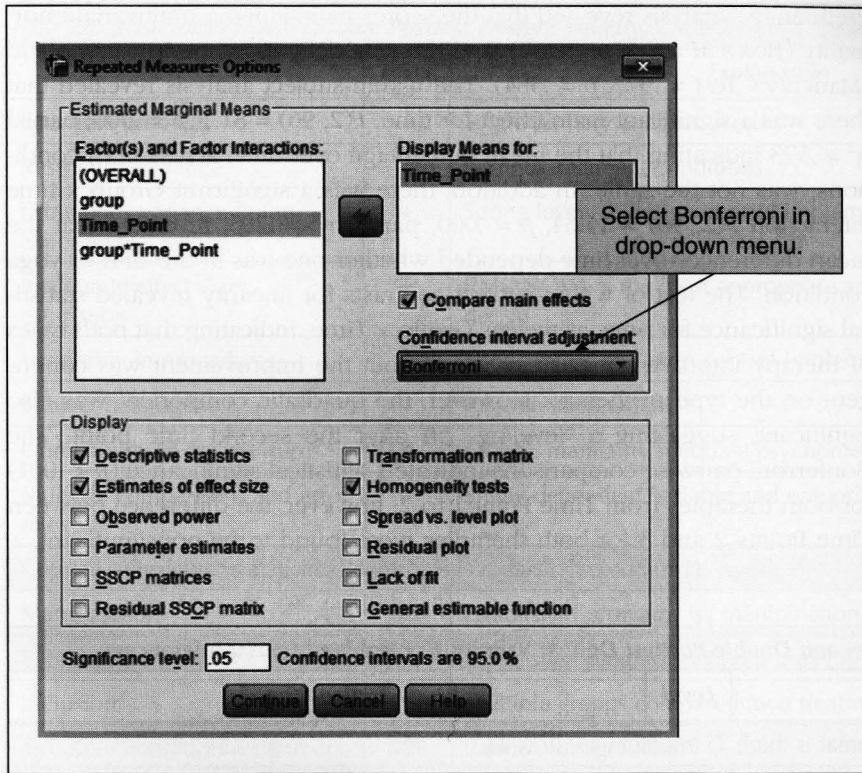
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	4384.807	1	4384.807	4827.310	.000	.990
group	45.927	1	45.927	50.561	.000	.513
Error	43.600	48	.908			



PAIRWISE COMPARISONS

Data → Split File





Note: Discard all other output except for the Pairwise Comparisons.

Pairwise Comparisons

Measure: OOLFS

group	(i) Time Point	(j) Time Point	Mean Difference (I - J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
						Lower Bound	Upper Bound
BFB Therapy	1	2	-3.000	.356	.000	-3.916	-2.084
		3	-3.240	.307	.000	-4.030	-2.450
	2	1	3.000	.356	.000	2.084	3.916
		3	-.240	.279	1.000	-.957	.477
	3	1	3.240	.307	.000	2.450	4.030
		2	.240	.279	1.000	-.477	.957
Yoga	1	2	-.880	.323	.036	-1.712	-.048
		3	-1.560	.317	.000	-2.375	-.745
	2	1	.880	.323	.036	.048	1.712
		3	-.880	.320	.132	-1.504	.144
	3	1	1.560	.317	.000	.745	2.375
		2	.880	.320	.132	-.144	1.504

Based on estimated marginal means

Results

An RM ANOVA was run to examine the differences and lasting effects between BFB therapy and a yoga program on perceived functionality.

Preliminary analysis revealed that the scores maintained a multivariate normality (Box's $M = 3.4, p = .787$), and the sphericity assumption was upheld (Mauchly's Test = .977, $p = .584$). The within-subject analysis revealed that there was a significant main effect for time, $F(2, 96) = 81.1, p = .000$, partial $\eta^2 = .573$ indicating that the mean percentage over time, across both conditions, was not the same. In addition, there was a significant Group \times time interaction $F(2, 96) = 15.64, p = .000$, partial $\eta^2 = .205$, revealing that the mean differences over time depended whether one was in the BFB or yoga condition. The test of within-subject contrasts for linearity revealed statistical significance for time, as well as Group \times Time, indicating that both types of therapy improved linearly over time, but the improvement was contingent on the type of therapy. However, the quadratic component was also significant, suggesting a "leveling" off after the second time point. The Bonferroni pairwise comparisons indicated statistical significance ($p < .001$) for both therapies from Time Point 1 to 2; however, the difference between Time Points 2 and 3 for both therapies were found to be nonsignificant.

2-Factor Pre- and Double Posttest Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Internal Validity —Threat is (high \uparrow) (medium \leftrightarrow) (low \downarrow)		
History	\uparrow	Inclusion of pretest
Maturation	\uparrow	Inclusion of pretest
Testing	\uparrow	Inclusion of pretest
Instrumentation	\downarrow	The instrument did not change between pre- and posttests
Statistical regression	\leftrightarrow	Random assignment applied
Attrition	\leftrightarrow	Length of time between measures
Selection bias	\downarrow	Random assignment applied
Combination of selection and other treatments	\downarrow	Random assignment and independent groups
Diffusion	\downarrow	Conditions are independent
Special treatment	\downarrow	Independent treatment conditions
Sequencing effects	n/a	Each group was exposed to only one condition
Statistical Conclusion Validity —Threat is (high \uparrow) (medium \leftrightarrow) (low \downarrow)		
Low statistical power	\downarrow	Adequate sample size
Assumption violation of statistical tests	\downarrow	Multivariate normality, sphericity, and independence confirmed

2-Factor Pre- and Double Posttest Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Error rate problem	↓	Single primary analysis used
Restriction of range	↓	Distinct treatment conditions
Extraneous variance in the experimental setting	↓	Strong levels of design control implemented
Inaccurate effect size estimation	↓	RM ANOVA is the most appropriate analysis for this design
Variability in procedures	↓	Strong levels of design control implemented
Subject heterogeneity	↓	Random assignment
Unreliability of the measures	↓	QOLFS maintains adequate psychometric properties
Multiple comparisons and error rates	↓	Single dependent variable and one primary analysis used
Control—Design and statistical control is (strong †††) (medium ††) (weak †)		
Manipulation	†††	Established programs w/ manipulation check
Elimination	†††	Established programs w/ manipulation check
Inclusion	†††	Multiple groups compared (two treatment conditions)
Group or condition assignment	†††	Random assignment
Statistical procedures	n/a	Data were not altered
Cause and Effect—Cause-effect determination is (strong †††) (medium ††) (weak †)		
Covariation	†††	Strong controls and minimal threats to internal and statistical conclusion validity
Temporal precedence	†††	Strong controls and minimal threats to internal and statistical conclusion validity
No plausible alternative explanations	††	Covariation and temporal precedence soundly confirmed, but replicable results should be obtained over time

The between-subject analysis revealed a statistically significant difference between the groups $F(1, 48) = 50.56, p = .000$. The effect size revealed there was a substantial difference between the groups (partial $\eta^2 = .513$), indicating the BFB therapy produced stronger effects. The estimated marginal mean illustrates the change over time and that the initial measurement (pretest) was similar for the BFB group ($M = 3.88, SD = 1.26$) and the yoga group ($M = 4.04, SD = .978$). However, the BFB group evidenced greater gains ($M = 6.88, SD = .97$) in the first posttest than the yoga group ($M = 4.92,$

$SD = 1.11$), and the differences were maintained in the follow-up test for the BFB group ($M = 7.12$, $SD = .97$) and the yoga group ($M = 5.60$, $SD = 1.08$).

Example E4 *3-Factor Crossover Design*

A military psychologist is interested in testing the effects of stress on reaction time in soldiers. The psychologist devised three different stress conditions (high, medium, and low) but only had access to six subjects. Therefore, each soldier will be exposed to each of the stress conditions and tested for simple reaction time (RT) following exposure to stress. The order of the conditions will be counterbalanced to control for sequencing effects. The high-stress condition will be labeled as *A*, the medium-stress condition will be labeled as *B*, and the low-stress condition will be labeled as *C*. All three conditions will follow an established standardized protocol for implementation and will use manipulation checks to ensure the stress conditions are implemented as intended to ensure the integrity of various design controls (e.g., manipulation and elimination).

General Research Question: What are the effects of three different conditions (A, B, and C) on outcome (DV: continuous data)?

Grouping Variable (3 levels): One IV at three levels—A (High), B (Medium), and C (Low)

DV: Continuous (reaction time)

<i>Subject</i>	<i>Condition</i>	<i>Test</i>	<i>Condition</i>	<i>Test</i>	<i>Condition</i>	<i>Test</i>
1	X_A	O_1	X_B	O_2	X_C	O_3
2	X_B	O_1	X_C	O_2	X_A	O_3
3	X_C	O_1	X_A	O_2	X_B	O_3
4	X_C	O_1	X_B	O_2	X_A	O_3
5	X_A	O_1	X_C	O_2	X_B	O_3
6	X_B	O_1	X_A	O_2	X_C	O_3

Time ►

Note: Any number of posttests and factors can be included, based on theoretical and logistical considerations.

Design: Experimental research using a within-subjects approach with a 3-factor crossover design

Null Hypothesis: The adjusted population means for all conditions are equal. In this example, $H_0: \mu_1 = \mu_2$ and μ_3 or the adjusted means of each condition are equivalent.

Statistic: One-Way Within-Subjects Analysis of Variance (RM ANOVA)

Statistical Power: Utilize G*Power Data Analysis

Specific Research Question: What are the effects of varying levels of stress on simple reaction time as measured in milliseconds?

Null Hypothesis: There is no difference in the reaction times (RT) between the high-, medium-, and low-stress conditions.

<i>Subject</i>	<i>Condition</i>	<i>Test</i>	<i>Condition</i>	<i>Test</i>	<i>Condition</i>	<i>Test</i>
1	High	RT	Medium	RT	Low	RT
2	Medium	RT	Low	RT	High	RT
3	Low	RT	High	RT	Medium	RT
4	Low	RT	Medium	RT	High	RT
5	High	RT	Low	RT	Medium	RT
6	Medium	RT	High	RT	Low	RT

Time ►

Note: The decision to consider when a subject is exposed to subsequent conditions should be considered, based on theoretical and logistical tenants (i.e., washout period).

DATA ENTRY

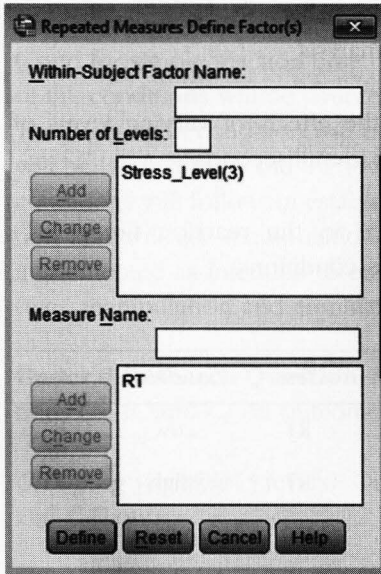
See data set for this particular example for an understanding on how to set up Variables in Variable view and how to enter data into SPSS for the within-subjects approach.

♦ COMMAND PROMPTS FOR SPSS

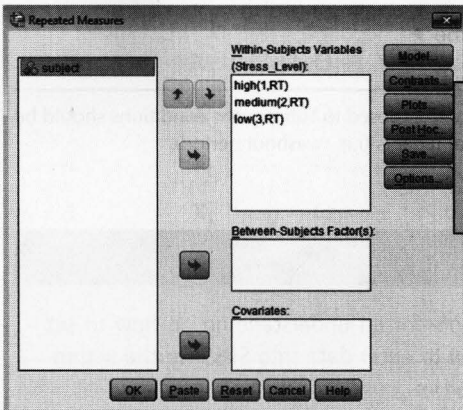
Data titled *3factorcrossover.sav*

RM ANOVA

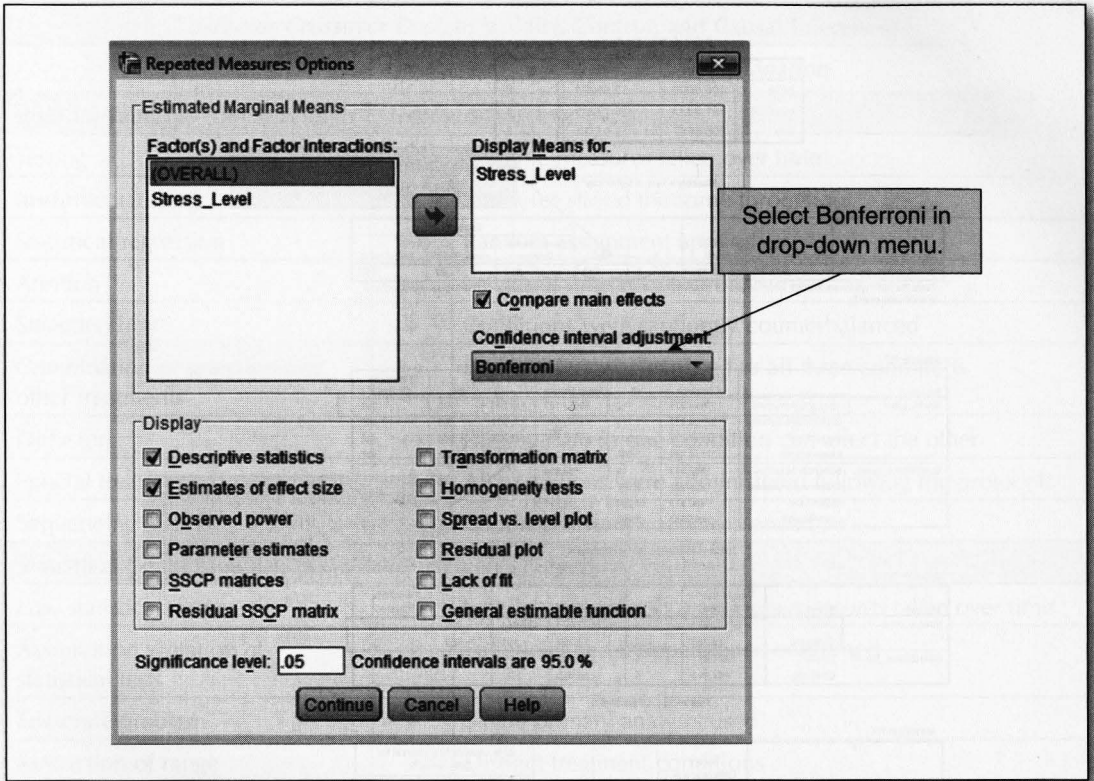
Analyze → General Linear Models → Repeated Measures



Enter the name of the WS factor as Stress-Level, and enter the number of time points; then select Add; enter the name of the DV in the Measure Name, and select Add; select Define.



Enter each stress level in order in the WS Variables box, and then select Options.



Results

An RM ANOVA was run to examine the differences between three different levels of stress (high, medium, and low), and simple reaction time was measured in milliseconds. Preliminary analysis revealed that the sphericity assumption was upheld (Mauchly's test = .732, $p = .693$). The within-subject analysis revealed that there was a significant effect, $F(2, 10) = 336.94$, $p = .000$, partial $\eta^2 = .985$, indicating that the mean differences across conditions were not the same. The test of within-subject contrasts for linearity revealed statistical significance ($p = .000$) for stress level, indicating that reaction time linearly increases from low to medium to the high condition. However, the quadratic component was not significant ($p = .763$), suggesting that the stress levels do not tend to level off between conditions. The Bonferroni pairwise comparisons indicated statistical significance ($p < .001$) between every stress level condition. The effect size revealed there was a strong effect between the conditions (partial $\eta^2 = .985$), indicating the amount of stress induced substantially impacted reaction time. The

Descriptive Statistics

	Mean	Std. Deviation	N
high	406.5000	13.61984	6
medium	301.8333	10.72225	6
low	200.6667	9.64711	6

Mauchly's Test of Sphericity^a

Mauchly's test is not significant, indicating sphericity.

Measure: RT

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Stress_Level	.833	.732	2	.693	.857	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

Tests of Within-Subjects Effects

Measure: RT

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Stress_Level	Sphericity Assumed	127114.333	2	63557.167	336.935	.000	.985
	Greenhouse-Geisser	127114.333	1.713	74189.540	336.935	.000	.985
	Huynh-Feldt	127114.333	2.000	63557.167	336.935	.000	.985
	Lower-bound	127114.333	1.000	127114.333	336.935	.000	.985
Error(Stress_Level)	Sphericity Assumed	1886.333	10	188.633			
	Greenhouse-Geisser	1886.333	8.567	220.189			
	Huynh-Feldt	1886.333	10.000	188.633			
	Lower-bound	1886.333	5.000	377.267			

Mauchly's test is not significant, therefore, the "Sphericity Assumed" row should be used.

Tests of Within-Subjects Contrasts

Measure: RT

Source	Stress_Level	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Stress_Level	Linear	127102.083	1	127102.083	495.944	.000	.990
	Quadratic	12.250	1	12.250	.101	.763	.020
Error(Stress_Level)	Linear	1281.417	5	256.283			
	Quadratic	604.917	5	120.983			

Pairwise Comparisons

Measure: RT

(I) Stress_Level	(J) Stress_Level	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
1	2	104.667	7.894	.000	76.769	132.564
	3	205.833 [*]	9.243	.000	173.169	238.498
2	1	-104.667	7.894	.000	-132.564	-76.769
	3	101.167 [*]	6.395	.000	78.566	123.767
3	1	-205.833 [*]	9.243	.000	-238.498	-173.169
	2	-101.167 [*]	6.395	.000	-123.767	-78.566

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

descriptive statistics further illustrate the point, revealing that the low-stress condition turned out the fastest reaction times ($M = 200.67, SD = 9.64$), while medium stress was still faster ($M = 301.83, SD = 10.72$) than the high-stress condition ($M = 406.50, SD = 13.61$).

3-Factor Crossover Design: Validity, Control, and Causal Inferences

	Level	Explanation
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
History	↑	Multiple measures taken over time
Maturation	↑	Multiple measures taken over time

3-Factor Crossover Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Testing	↑	Multiple measures taken over time
Instrumentation	↓	Measures stayed the same throughout
Statistical regression	↔	Random assignment applied
Attrition	↔	Length of time between measures
Selection bias	↓	Conditions were randomly counterbalanced
Combination of selection and other treatments	↔	Each subject was exposed to all three conditions
Diffusion	↔	Participation in one condition can affect the other
Special treatment	↓	All conditions were administered following the protocols
Sequencing effects	↔	Conditions were counterbalanced
Statistical Conclusion Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Low statistical power	↔	Small sample size but multiple measures taken over time
Assumption violation of statistical tests	↓	Sphericity and independence confirmed
Error rate problem	↓	Single primary analysis used
Restriction of range	↓	Distinct treatment conditions
Extraneous variance in the experimental setting	↓	Strong levels of design control implemented
Inaccurate effect size estimation	↓	RM ANOVA is the most appropriate analysis for this design
Variability in procedures	↓	Strong levels of design control implemented
Subject heterogeneity	↔	Subjects share likenesses
Unreliability of the measures	↓	Simple reaction time measured electronically w/o error
Multiple comparisons and error rates	↓	Single dependent variable and one primary analysis used
Control —Design and statistical control is (strong †††) (medium ††) (weak †)		
Manipulation	†††	Established protocol w/ manipulation check
Elimination	†††	Established protocol w/ manipulation check
Inclusion	†††	Three conditions included
Group or condition assignment	†††	Counterbalanced conditions
Statistical procedures	n/a	Data were not altered

(Continued)

(Continued)

<i>3-Factor Crossover Design: Validity, Control, and Causal Inferences</i>		
	<i>Level</i>	<i>Explanation</i>
Cause and Effect —Cause-effect determination is (strong TTT) (medium TT) (weak T)		
Covariation	TT	Strong controls and medium threats to internal and statistical conclusion validity
Temporal precedence	TT	Strong controls and medium threats to internal and statistical conclusion validity
No plausible alternative explanations	TT	Covariation and temporal precedence confirmed, but replicable results should be obtained over time

Example E5 3×3 Latin-Square Design

A sport psychologist is interested in testing the performance of different types of golf putters while controlling for varying levels of putt difficulty and anxiety. He will carry out this experiment by designing a basic 3×3 Latin-square design. The design will include the primary factor, which is putter type. The three types of putters are blade, mallet, and a futuristic. Manipulation will be the design control used for the two nuisance (blocking) factors of anxiety and putt difficulty; each will be separated into three levels. A high, medium, and low anxiety condition will be combined with easy, medium, and hard putting conditions while testing one of the three different types of putters. Performance will be measured as the radial distance the ball lands from the hole and will be measured in centimeters. A total of 27 high-level golfers will be identified and then will be randomly assigned to one of the three conditions.

General Research Question: What are the effects of primary factory (IV: categorical at p levels) on outcome (DV: continuous data) while controlling for blocking factor 1 (IV: categorical at p levels) and blocking factor 2 (IV: categorical at p levels)?

Grouping Variable: Three IVs each with three levels (2 blocking factors and 1 primary factor); $k = 3$ factors; L_1 (Block) = 3 levels; L_2 (Block) = 3 levels; L_3 (Primary) = 3 levels

DV: Continuous (centimeters [radial distance from the hole])

3 × 3 Latin-Square Design

Group	Blocking Factor 1	Blocking Factor 2		
		Level 1	Level 2	Level 3
1	Level 1	A	B	C
2	Level 2	C	A	B
3	Level 3	B	C	A

Note: An important design assumption for the Latin-square design is that there is no interaction between the procedural factors as well as the primary factor because not all combinations of the levels of each factor are tested (i.e., interaction effects cannot be analyzed); however, interactions commonly occur between psychologically related constructs, which explains why this design is typically not applied in education and the social and behavioral sciences.

Design: Experimental research using a between-subjects approach with a 3 × 3 Latin-square design

Null Hypothesis: The adjusted population means for all groups are equal. In this example, $H_0: t_1 = t_2 = \dots = t_p = 0$ or $H_0: T_t = 0$.

Statistic: General Linear Models One-Way Between-Subjects ANOVA

Statistical Power: The number of levels of the two blocking factors is multiplied to determine the number of runs or number of participants required for each condition.

Specific Research Question: What are the effects of golf putter type on performance when controlling for levels of anxiety and putt difficulty?

Null Hypothesis: There is no difference in performance between the different levels of putt difficulty. There is no difference in performance between the different levels of anxiety. There is no difference in performance between the different types of putters.

$k = 3$ factors (2 blocking factors and 1 primary factor)

L_1 (Block) = Anxiety (3 levels)—High, Medium, Low

L_2 (Block) = Putt Difficulty (3 levels)—Easy, Medium, Hard

L_3 (Primary) = Putter Type (3 levels)—Blade, Mallet, Futuristic

$n = 9$ subjects per condition (number of levels of the 2 blocking factors are multiplied to determine the number of runs or number of participants, $3_1 * 3_2 = 9$)

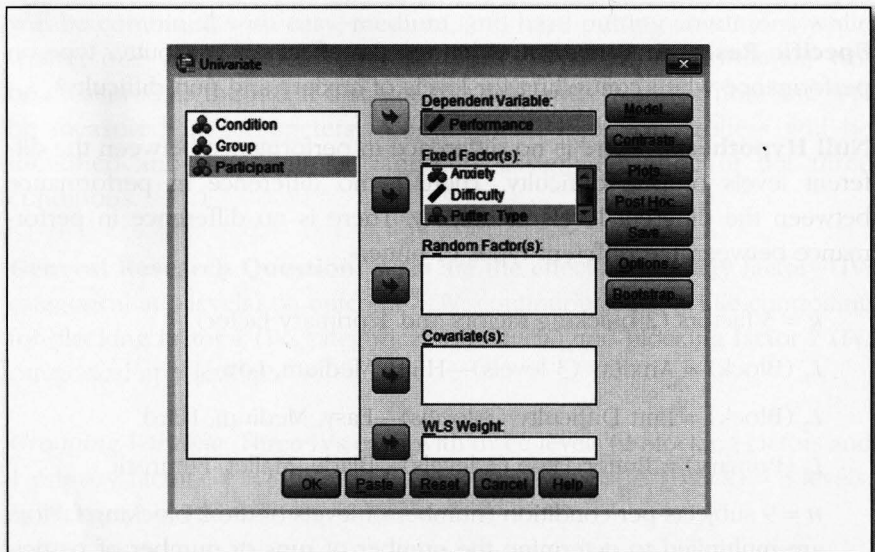
Group	Anxiety	Putt Difficulty		
		Easy	Medium	Hard
1 (n = 9)	High	A (Blade)	B (Mallet)	C (Futuristic)
2 (n = 9)	Medium	C (Futuristic)	A (Blade)	B (Mallet)
3 (n = 9)	Low	B (Mallet)	C (Futuristic)	A (Blade)

Note: Sequencing effects are a major threat to internal validity with this design. Therefore, based on theoretical and logistical considerations, a certain amount of time should elapse between conditions (i.e., a “washout” period). Also, not every combination of levels for putt difficulty, anxiety, and type of putter are tested; hence, this design is referred to as a *fractional factorial* design. A complete factorial design would be a 3 x 3 x 3 design to correct for this; however, it would require 27 cells (as opposed to the current nine cells), which may not be feasible due to cost, time, and access to participants. Another alternative to testing all combinations is to run a repeated Latin square to include all combinations of levels for each factor and is presented in Appendix A.

DATA ENTRY

See data set for this particular example for an understanding on how to set up Variables in Variable view and how to enter data into SPSS for the Latin-square design.

◆ COMMAND PROMPTS FOR SPSS

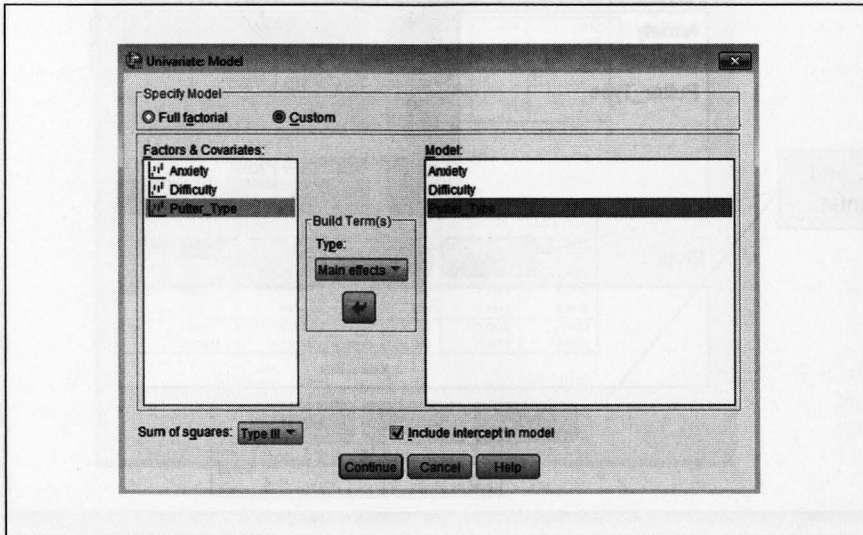


Note: Select **Options**, then *Estimates of Effect Size and Descriptives*.

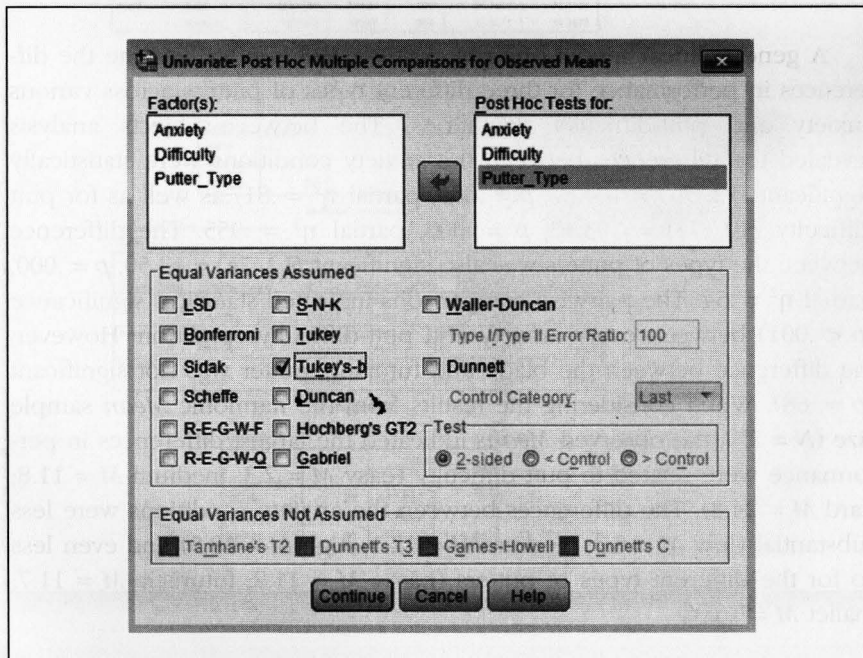
Data titled *3x3Latinsquare.sav*

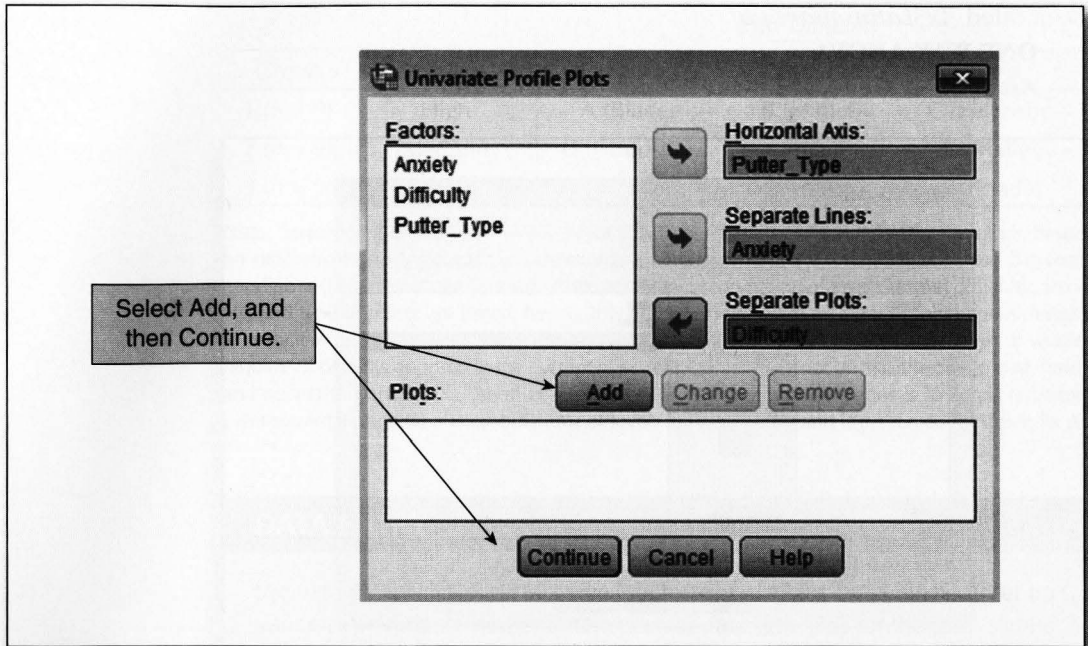
ONE-WAY ANOVA

Analyze → General Linear Model → Univariate



Note: Select **Main effects** under Build Term(s).





Results

A general linear model one-way ANOVA was run to examine the differences in performance for three different types of putters across various anxiety and putt-difficulty conditions. The between-subjects analysis revealed the differences between the anxiety conditions were statistically significant $F(2, 74) = 164.87, p = .000$, partial $\eta^2 = .817$, as well as for putt difficulty $F(2, 74) = 793.17, p = .000$, partial $\eta^2 = .955$. The difference between the types of putters was also significant $F(2, 74) = 43.50, p = .000$, partial $\eta^2 = .54$. The pairwise comparisons indicated statistical significance ($p < .001$) between every anxiety and putt-difficulty condition. However, the difference between the blade and futuristic putter was not significant ($p = .68$). When considering the results from the harmonic *Mean* sample size ($N = 27$), the observed *Means* indicated the largest differences in performance were related to putt difficulty (easy $M = 2.3$; medium $M = 11.8$; hard $M = 24.8$). The differences between the anxiety conditions were less substantial (low $M = 8.3$; medium $M = 12.1$; high $M = 18.5$) and even less so for the different types of putters (blade $M = 11.2$; futuristic $M = 11.7$; mallet $M = 16.0$).

Tests of Between-Subjects Effects

Dependent Variable: Performance

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	8742.815 ^a	8	1457.136	333.846	.000	.964
Intercept	13585.198	1	13585.198	3112.517	.000	.977
Anxiety	1439.210	2	719.605	164.969	.000	.817
Difficulty	6923.877	2	3461.938	793.188	.000	.955
Putter_Type	379.728	2	189.864	43.500	.000	.540
Error	322.988	74	4.365			
Total	22651.000	81				
Corrected Total	9065.802	80				

a. R Squared = .964 (Adjusted R Squared = .961)

Multiple Comparisons

Dependent Variable: Performance

Tukey HSD

(I) Anxiety	(J) Anxiety	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
High	Medium	6.3704 [*]	.56860	.000	5.0104	7.7303
	Low	10.2222 [*]	.56860	.000	8.8623	11.5822
Medium	High	-6.3704 [*]	.56860	.000	-7.7303	-5.0104
	Low	3.8519 [*]	.56860	.000	2.4919	5.2118
Low	High	-10.2222 [*]	.56860	.000	-11.5822	-8.8623
	Medium	-3.8519 [*]	.56860	.000	-5.2118	-2.4919

Performance

Tukey B^{a,b}

Anxiety	N	Subset		
		1	2	3
Low	27	9.2593		
Medium	27		12.1111	
High	27			18.4815

Multiple Comparisons

Dependent Variable: Performance

Tukey HSD

(I) Difficulty	(J) Difficulty	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Easy	Medium	-9.5185 [*]	.56860	.000	-10.8785	-8.1586
	Hard	-22.5556 [*]	.56860	.000	-23.9155	-21.1956
Medium	Easy	9.5185 [*]	.56860	.000	8.1586	10.8785
	Hard	-13.0370 [*]	.56860	.000	-14.3970	-11.6771
Hard	Easy	22.5556 [*]	.56860	.000	21.1956	23.9155
	Medium	13.0370 [*]	.56860	.000	11.6771	14.3970

Performance

Tukey B^{a,b}

Difficulty	N	Subset		
		1	2	3
Easy	27	2.2593		
Medium	27		11.7778	
Hard	27			24.8148

Differences between the blade and futuristic were not significant.

Multiple Comparisons

Dependent Variable: Performance

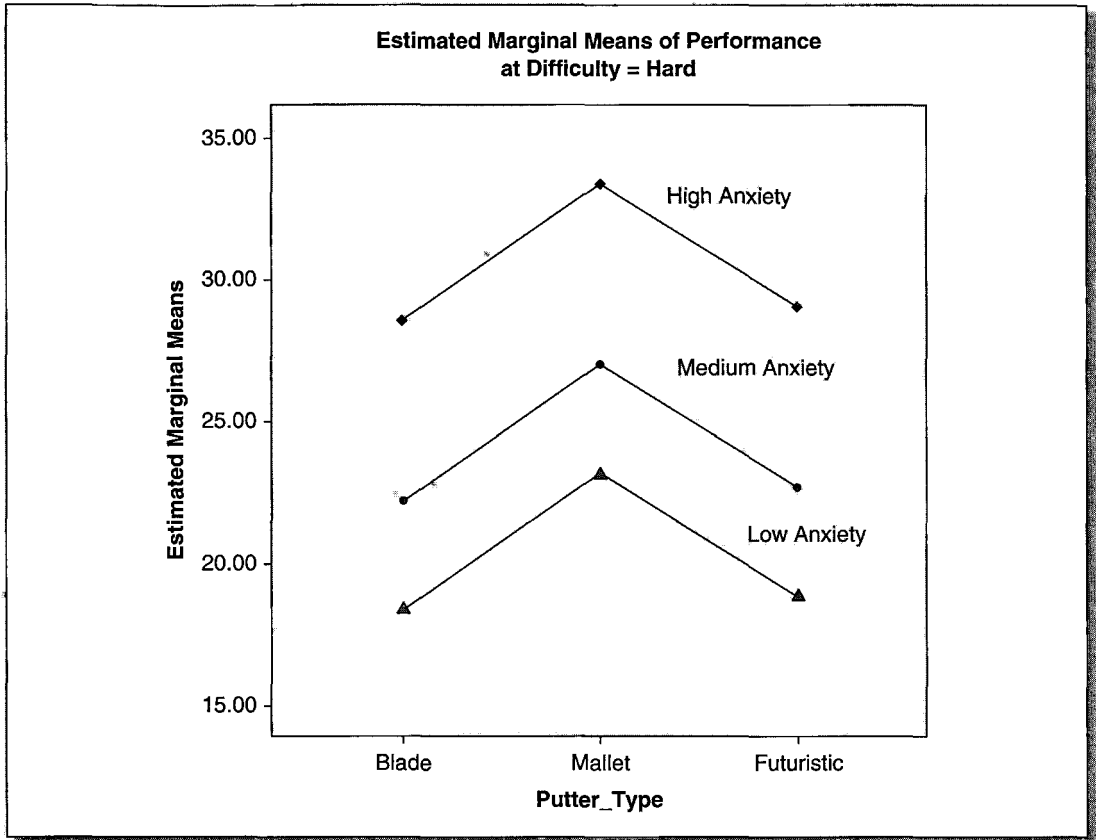
Tukey HSD

(I) Putter_Type	(J) Putter_Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Blade	Mallet	-4.8148 [*]	.56860	.000	-6.1748	-3.4548
	Futuristic	-.4815	.56860	.675	-1.8414	.8785
Mallet	Blade	4.8148 [*]	.56860	.000	3.4548	6.1748
	Futuristic	4.3333 [*]	.56860	.000	2.9734	5.6933
Futuristic	Blade	.4815	.56860	.675	-.8785	1.8414
	Mallet	-4.3333 [*]	.56860	.000	-5.6933	-2.9734

Performance

Tukey B^{a,b}

Putter_Type	N	Subset	
		1	2
Blade	27	11.1852	
Futuristic	27	11.6667	
Mallet	27		16.0000



Difficulty	Anxiety	Putter_Type	Mean	Std. Deviation
Easy	High	Blade	3.5556	.88192
		Total	3.5556	.88192
	Medium	Futuristic	1.6667	.70711
		Total	1.6667	.70711
	Low	Mallet	1.5556	.52705
		Total	1.5556	.52705
Medium	High	Mallet	21.8889	1.16667
		Total	21.8889	1.16667
	Medium	Blade	10.1111	.60093
		Total	10.1111	.60093
	Low	Futuristic	3.3333	.50000
		Total	3.3333	.50000
Hard	High	Futuristic	30.0000	1.41421
		Total	30.0000	1.41421
	Medium	Mallet	24.5556	1.13039
		Total	24.5556	1.13039
	Low	Blade	19.8889	1.45297
		Total	19.8889	1.45297

Latin-Square Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
History	↑	Multiple measures taken over time
Maturation	↑	Multiple measures taken over time
Testing	↑	Multiple measures taken over time
Instrumentation	↓	Measures stayed the same throughout
Statistical regression	↔	Random assignment applied
Attrition	↔	Length of time between measures
Selection bias	↓	Conditions were randomly counterbalanced
Combination of selection and other treatments	↓	Each subject was exposed to all three conditions
Diffusion	↔	Participation in one condition can affect the other
Special treatment	↓	All conditions were administered following the protocols
Sequencing effects	↑	Conditions were counterbalanced but may affect performance in the subsequent conditions
Statistical Conclusion Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Low statistical power	↓	Adequate sample size
Assumption violation of statistical tests	↔	Data were normally distributed, but interaction effects may be present between variables but were not tested
Error rate problem	↓	Single primary analysis used
Restriction of range	↓	Distinct treatment conditions resulted in wide ranges in performance
Extraneous variance in the experimental setting	↓	Strong levels of design control implemented
Inaccurate effect size estimation	↓	ANOVA is the most appropriate analysis for this design
Variability in procedures	↓	Strong levels of design control implemented
Subject heterogeneity	↔	Subjects share likenesses
Unreliability of the measures	↓	Radial distance measured in centimeters
Multiple comparisons and error rates	↓	Single dependent variable and one primary analysis used

(Continued)

(Continued)

<i>Latin-Square Design: Validity, Control, and Causal Inferences</i>		
	<i>Level</i>	<i>Explanation</i>
Control —Design and statistical control is (strong †††) (medium ††) (weak †)		
Manipulation	†††	Established protocol w/ manipulation check
Elimination	†††	Established protocol w/ manipulation check
Inclusion	†††	Multiple conditions included
Group or condition assignment	†††	Counterbalanced conditions
Statistical procedures	n/a	Data were not altered
Cause and Effect —Cause-effect determination is (strong †††) (medium ††) (weak †)		
Covariation	††	Strong controls and medium threats to internal and statistical conclusion validity
Temporal precedence	††	Strong controls and medium threats to internal and statistical conclusion validity
No plausible alternative explanations	††	Covariation and temporal precedence confirmed, but replicable results should be obtained over time

Example E6 *Wait-List Continuation Design (for Random Assignment)*

A pharmacology researcher is interested in the effects of omega-3 oil (fish oil) on cholesterol. According to the literature, omega-3 oil has a half-life (the time it takes to reduce initial concentration by 50%) of approximately 70 hours. He recruits 32 individuals from local clinics who were prescreened to have low-density lipoprotein (LDL) levels of 190 (very high) and above. LDL is considered bad cholesterol and serves as a good predictor of heart disease and stroke. Participants will be randomly assigned to Group 1 ($n = 16$) or Group 2 ($n = 16$). On the first day, Group 1 participants will have their LDL tested via a lipid panel (O_1 and O_3). Group 1 will receive a 3-month supply of omega-3 oil, and participants will be trained how to use the supplement daily. Group 2 will receive a 3-month supply of a placebo (sugar pills). At the 3-month mark, all participants will be tested via lipid panel (O_1 and O_4) and ***if the treatment is effective***, the conditions will be reversed. Group 2 will then receive a 3-month supply of omega-3 oil,

while Group 1 will receive a 3-month supply of the placebo pill. At the 6-month mark, all participants will complete the lipid panel test (O_2 and O_5).

General Research Question: To what extent do those who receive the treatment (Group 1) differ from those who do not receive treatment (Group 2) on their outcome (DV: continuous data), as well as the impact of the treatment over time?

Grouping Variable (2 levels): One IV at two levels = Group 1 receives treatment (omega-3 oil) and Group 2 does not receive treatment

DV: Continuous (LDL)

Group	Treatment	Test	Treatment	Test	–	Test
1	X_A	O_1	– Time delay	O_2	–	–
2	–	O_3	X_A	O_4	– Time delay	O_5

Time ►

Note: This is a design variant to the switching-replication design when random assignment can be used. The time delay is based on theoretical and logistical considerations. Group 1 is not affected by history or maturation because they do not receive the pretest. This is a good design variant when it is considered unethical to withhold treatment from one group.

Design: Experimental research using a between-subjects approach with a wait-list continuation design

Null Hypothesis: The adjusted population means for all groups are equal.
 $H_0: \mu_1 = \mu_3$ and $\mu_1 + \mu_4 = \mu_2 + \mu_5$.

Statistic: Independent samples *t* test for Time Points O_1 and O_3 ; a dependent or paired samples *t* test to examine the lasting effects for $O_1 + O_4$ compared to $O_2 + O_5$.

Statistical Power: Utilize G*Power Data Analysis

Research Questions: To what extent does omega-3 oil impact the LDL levels of individuals at risk of heart disease and stroke? Is there a lasting effect of the treatment on LDL levels?

Null Hypothesis: Participants who use omega-3 oil will demonstrate no change in their LDL levels across time. The group that receives the treatment first (Group 1) will not differ from the group that receives no treatment first (Group 2) on levels of LDL.

Group	Assignment	Treatment	Test1	Treatment	Test2	–	Test3
1 (n = 16)	R	Omega-3 oil	LDL	Time delay	LDL	–	–
2 (n = 16)	R	Placebo	LDL	Omega-3 oil	LDL	Time delay	LDL

Sample Data Collection Spreadsheet

Participant	Group 1		Participant	Group 2		
	1	2		1	2	3
1	207	241	17	249	243	241
2	182	221	18	220	176	239
3	272	241	19	261	179	223
.			.			
.			.			
16	210	246	32	247	192	261

Note: Data represented as mean scores on LDL for 32 participants (16 for Group 1 and 16 for Group 2).

DATA ENTRY

See data set for this particular example for an understanding on how to set up Variables in Variable view and how to enter data into SPSS for the between-subjects approach and pre- and posttest designs.

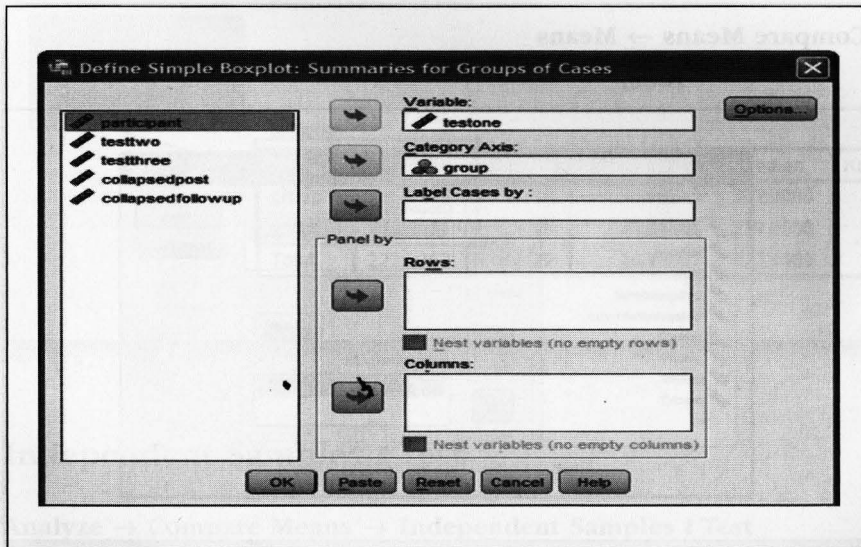
	group	participant	testone	testtwo	testthree	collapsedpost	collapsedfollowup
1	1.00	1.00	207.00	241.00	.	207.00	241.00
2	1.00	2.00	182.00	221.00	.	182.00	221.00
3	1.00	3.00	272.00	241.00	.	272.00	241.00
30	2.00	64.00	223.00	203.00	189.00	208.00	189.00
31	2.00	65.00	197.00	184.00	247.00	203.00	247.00
32	2.00	66.00	247.00	192.00	261.00	184.00	261.00

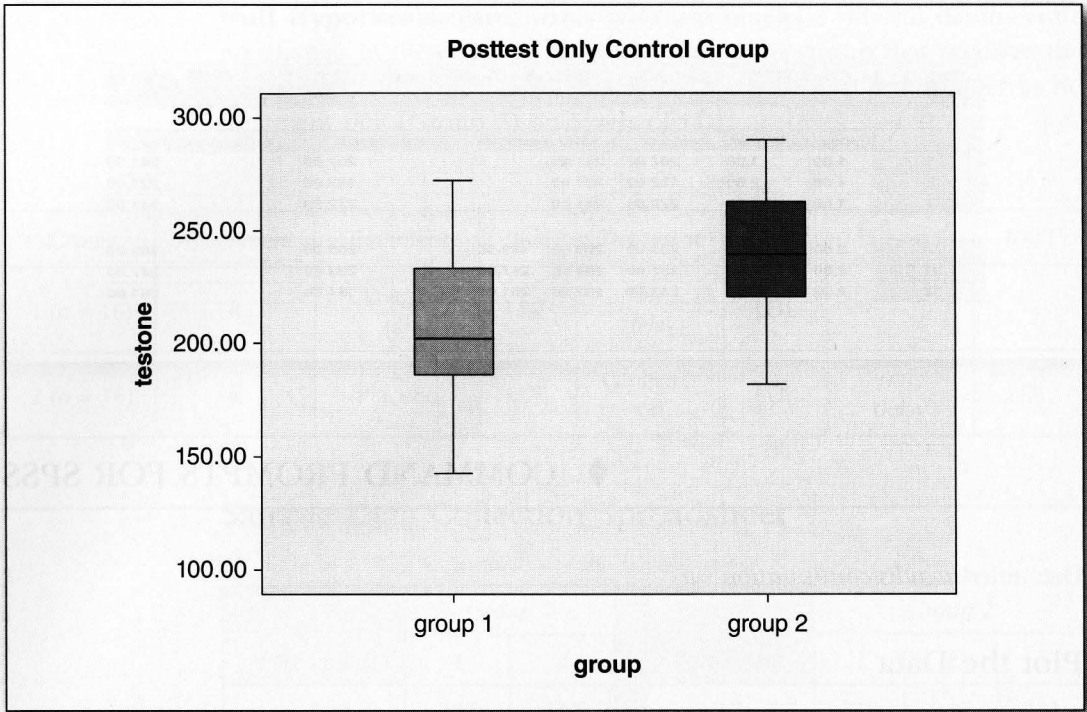
♦ COMMAND PROMPTS FOR SPSS

Data titled *waitlistcontinuation.sav*

Plot the Data

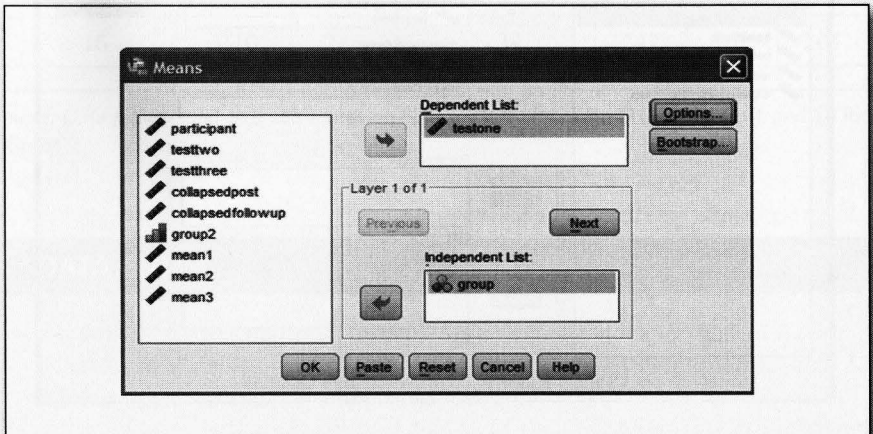
Graphs → Legacy Dialogs → Boxplot → Simple

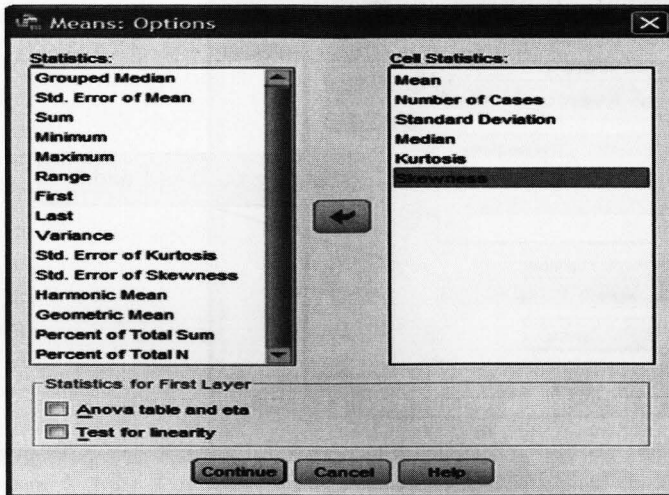




Descriptives

Compare Means → Means





As a general rule, Skewness and Kurtosis within +1 to -1 indicates that the data is normally distributed.

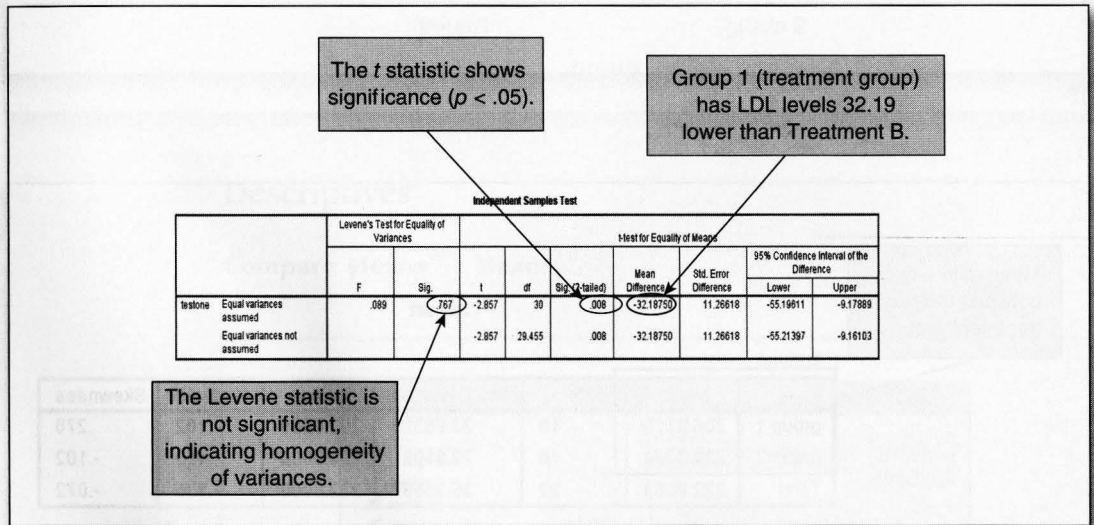
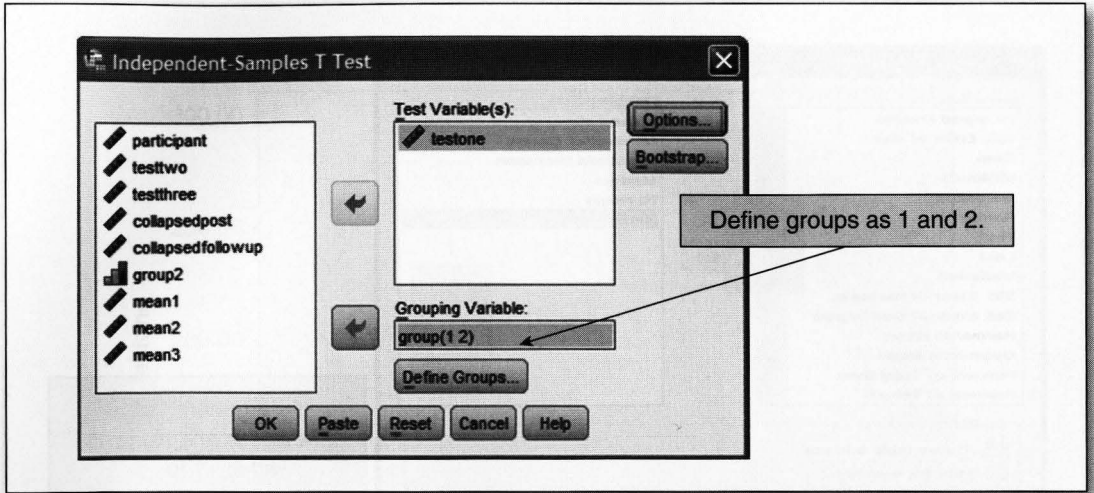
Mean gain scores between the two treatment groups

Report

testone							
group	Mean	N	Std. Deviation	Median	Kurtosis	Skewness	
group 1	206.8125	16	33.96315	202.0000	-.102	.276	
group 2	239.0000	16	29.61981	239.0000	-.473	-.102	
Total	222.9063	32	35.35567	222.0000	-.525	-.072	

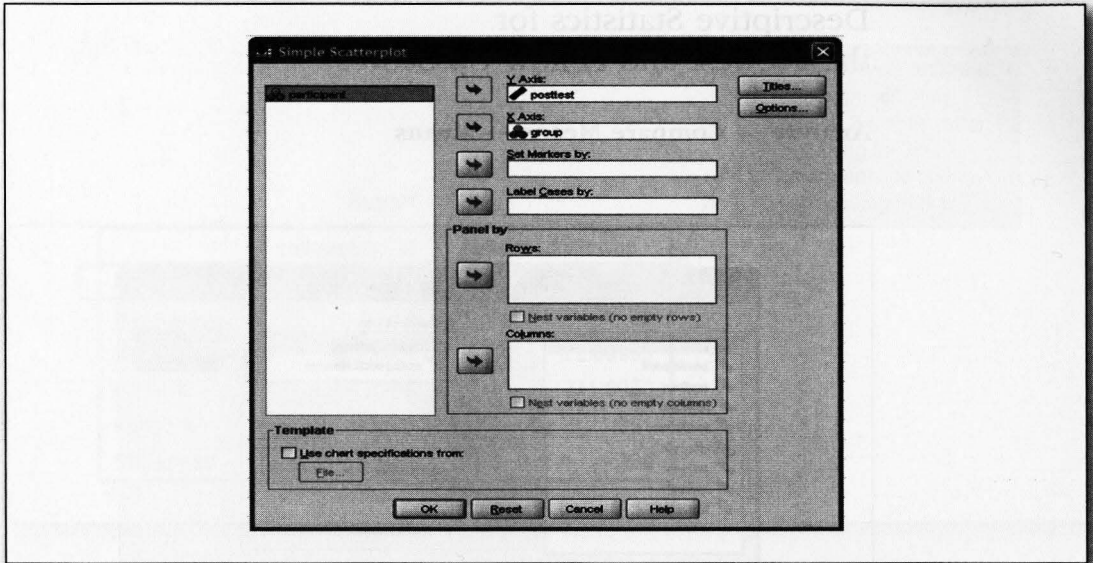
Independent Samples *t* Test

Analyze → Compare Means → Independent Samples *t* Test

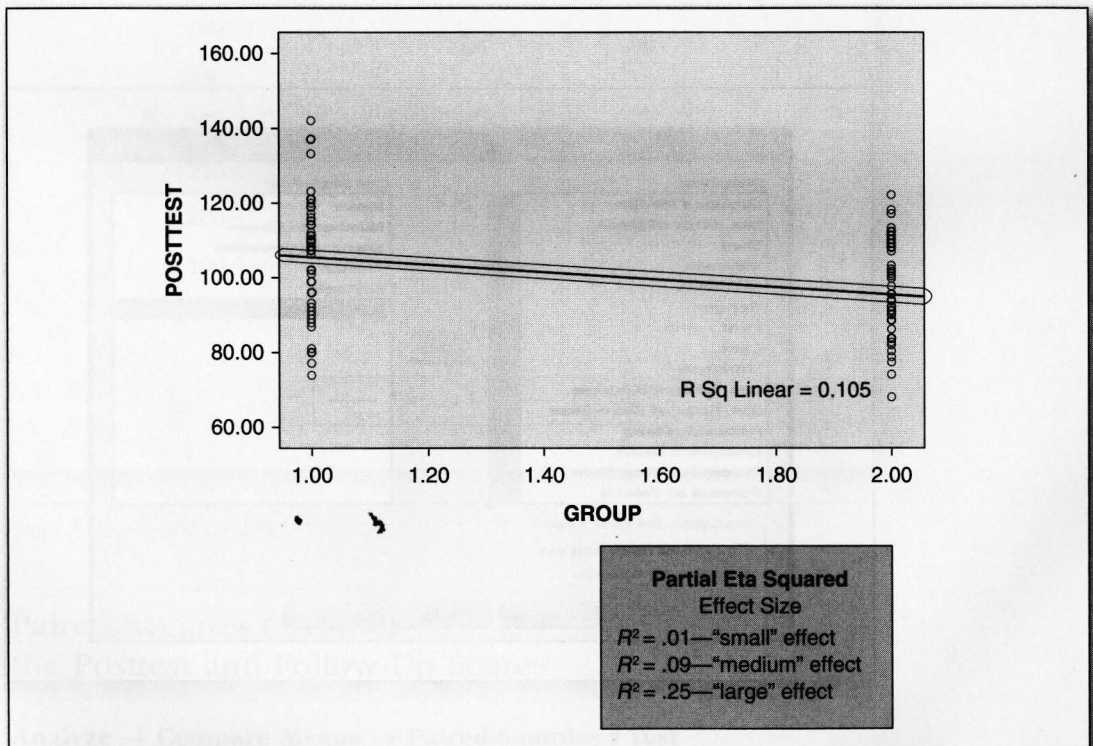


Point Biserial Correlation Coefficient (Effect Size)

Graph → Legacy Dialogues Test → Scatter/Dot → Simple Scatter



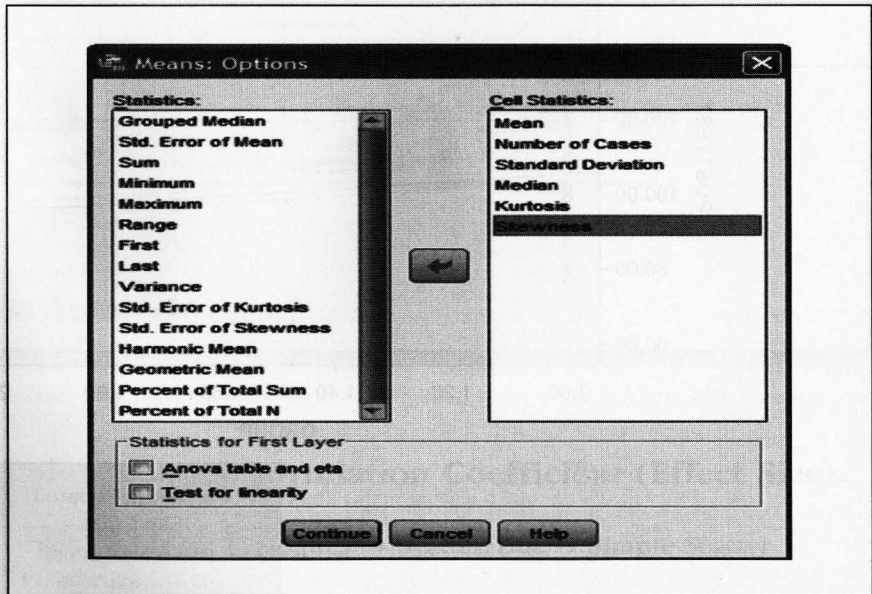
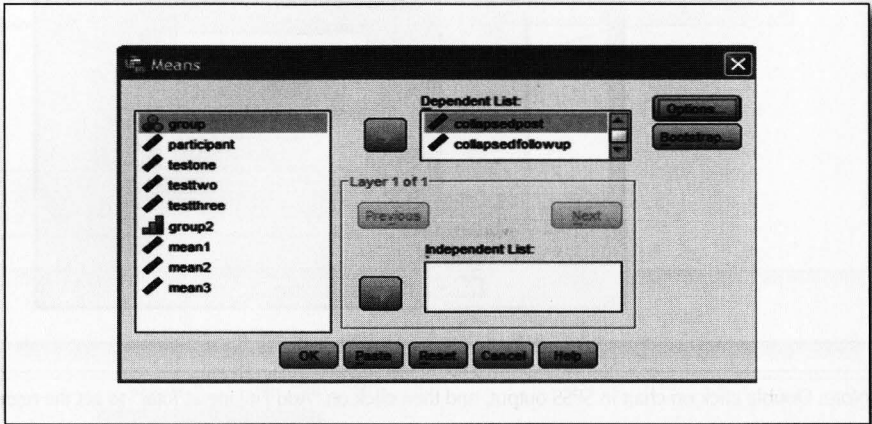
Note: Double click on chart in SPSS output, and then click on "Add Fit Line at Total" to get the regression line and R^2 value.



Note: R^2 value of .105.

Descriptive Statistics for the Posttest and Follow-Up Scores

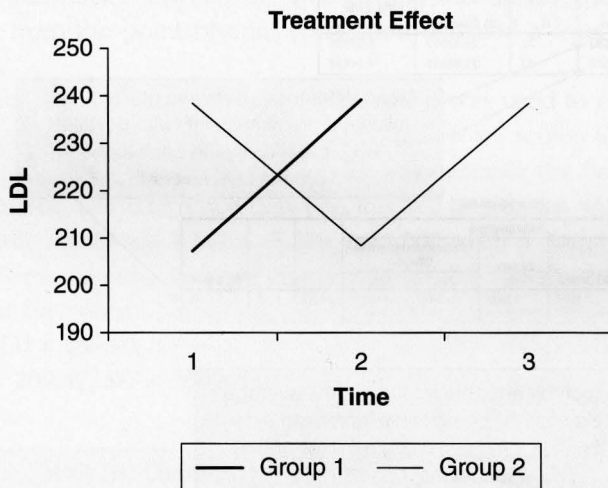
Analyze → Compare Means → Means



Report

	collapsedpost	collapsedfollowup
Mean	209.4687	238.4375
N	32	32
Std. Deviation	28.33553	22.88955
Median	208.0000	241.0000
Kurtosis	.001	.091
Skewness	.072	-.360

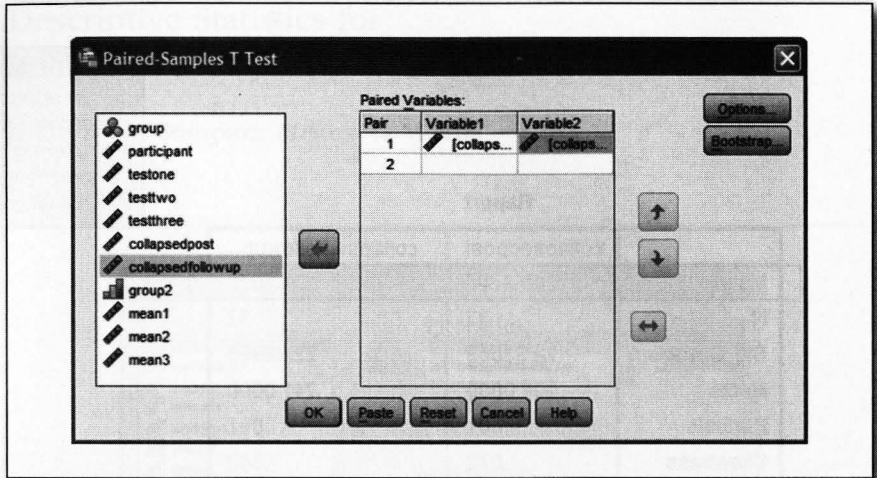
The LDL levels returned to approximately baseline 3 months after taking the last treatment, indicating no lasting effect.



Note: Time 1 is posttest data.

Paired Samples *t* Test for the Posttest and Follow-Up Scores

Analyze → Compare Means → Paired-Samples *t* Test



Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	collapsedpost	209.4688	32	28.33553	5.00906
	collapsedfollowup	236.4375	32	22.88955	4.04634

Mean differences between posttest and follow-up are significant (with posttest LDL levels dropping back to the pretest LDL levels).

Paired Samples Test

		Paired Differences			95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	collapsedpost-collapsedfollowup	-28.96875	31.41757	5.55389	-40.29598	-17.64151	-5.216	31	.000

Three months after the treatment, the average LDL levels of all 32 participants increased by ~ 29 points, indicating no lasting benefits of the treatment.

Effect Size

Converting *t*-value into *r*-value

$$r = \frac{\sqrt{t^2}}{t^2 + df}$$

$$r = \frac{\sqrt{-5.55^2}}{-5.55^2 + 31}$$

$$r = \frac{\sqrt{30.8}}{61.8} \oplus .71$$

Coefficient *r* Effect Size

r = .10—"small" effect: effect explains 1% of total variance
r = .30—"medium" effect: effect explains 9% of the total variance
r = .50—"large" effect: effect explains 25% of the variance

Results

An independent samples *t* test was used to compare the gain scores on the levels of an LDL test for participants who received omega-3 oils and those who received the placebo (i.e., Time Points O₁ and O₃). The difference between the two conditions was found to be statistically significant—*t*(30) = -2.86, *p* < .008. These findings indicate that individuals who received the omega-3 oils on average had lower levels of LDL (*M* = 206.8, *SD* = 33.96) than those who received the placebo (*M* = 239.0, *SD* = 29.61). The *Mean* difference between the two groups was 32.19. The effect size calculation from the point biserial correlation revealed a "medium" effect (*R*² = .105).

A paired samples *t* test (dependent samples) was used to compare the combined posttest scores and the combined follow-up scores on the levels of LDL (i.e., O₁ + O₄ compared to O₂ + O₅) to examine the lasting effects of the treatment. The difference between the two conditions were found to be statistically significant *t*(31) = -5.216, *p* < .000, with a "large" effect size (*r* = .71). These findings indicate that, on average, 3 months following the treatment of the omega-3 oils, levels of LDL returned close to original baseline levels (*M* = 238.43, *SD* = 22.89), compared to the immediate follow-up levels (*M* = 209.47, *SD* = 28.33).

Wait-List Continuation Design: Validity, Control, and Causal Inferences

	Level	Explanation
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
History	↑	Inclusion of multiple tests, including a pretest for Group 2
Maturation	↑	Inclusion of multiple tests, including a pretest for Group 2
Testing	↑	Inclusion of multiple tests, including a pretest for Group 2
Instrumentation	↓	Assessment of LDLs is stable

(Continued)

(Continued)

Wait-List Continuation Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Statistical regression	↔	Random assignment applied and time delay
Attrition	↔	Length of time between measures, including time delay
Selection bias	↓	Random assignment applied
Combination of selection and other treatments	↓	Random assignment and independent groups
Diffusion	↓	Conditions are independent but both groups receive treatment
Special treatment	↓	Independent treatment conditions
Sequencing effects	↓	Conditions are independent but both groups receive treatment
Statistical Conclusion Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Low statistical power	↓	Adequate sample size
Assumption violation of statistical tests	↓	Normality of data assessed
Error rate problem	↔	Combination of data and multiple types of analysis
Restriction of range	↓	One type of treatment
Extraneous variance in the experimental setting	↓	Strong levels of design control implemented
Inaccurate effect size estimation	↓	<i>t</i> tests are robust analyses
Variability in procedures	↓	Strong levels of design control implemented
Subject heterogeneity	↓	Random assignment
Unreliability of the measures	↓	LDL level assessment is standard
Multiple comparisons and error rates	↓	Single dependent variable
Control —Design and statistical control is (strong †††) (medium ††) (weak †)		
Manipulation	†††	Controlled for patient diet and activity level
Elimination	†††	Controlled for patient diet and activity level
Inclusion	†††	Multiple groups compared (two treatment conditions)
Group or condition assignment	†††	Random assignment
Statistical procedures	††	Data collapsed to examine lasting effects

Wait-List Continuation Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Cause and Effect —Cause-effect determination is (strong ttt) (medium tt) (weak t)		
Covariation	ttt	Strong controls and minimal threats to internal and statistical conclusion validity
Temporal precedence	ttt	Strong controls and minimal threats to internal and statistical conclusion validity
No plausible alternative explanations	tt	Covariation and temporal precedence soundly confirmed, but replicable results should be obtained over time

Example E7 *Regression-Discontinuity (Pretest-Posttest Control Group) Design*

A principal of a charter school understands that she has many students who performed poorly on last year's state math achievement test (SMAT). She is concerned that they will not do well on the upcoming SMAT, which will likely result in the school getting negative attention from the community. She plans to organize a remedial math program for all the students who meet the criteria and then test the results of this math program. Not all students need remediation; therefore, she'll have to determine which students need it and then find a way to compare their improvements to the improvement of those who don't go through the remedial program. Considering the circumstances, a regression-discontinuity approach will most likely reveal the results that she is looking for. In determining the cutoff score, theoretical considerations are used, which then leads to a decision of which participants are exposed to treatment and which are assigned to the control group. The most appropriate application would be to decide on a cutoff score that would allow for the assignment of participants who are in most "need" of the treatment. This need is determined via the distribution of scores on the pretest. In this particular example, the range on the SMAT is 0–28. Individuals who scored from 0 to 7 on the SMAT will be considered to be "in need" of supplementary instruction. The descriptive stats on the pretest indicated the actual pretest cutoff score was 6.8. Therefore, all students who scored a 6.8 or less will be assigned to the treatment condition. All others will be factored into the control condition.

Research Question: To what extent do those who are in most need of treatment and receive it (Group 1) differ from those who do not receive treatment (Group 2) on their outcome (DV: continuous data)?

Grouping Variable (2 levels): One IV and one control group = Group 1 receives treatment (remedial math program), and Group 2 receives no treatment

DV: Continuous (SMAT)

ANCOVA-SPECIFIC ASSUMPTIONS

1. A linear relationship exists between the posttest and the pretest (i.e., homogeneity of regression assumption).
2. The covariate is measured without error.

REGRESSION DISCONTINUITY-SPECIFIC ASSUMPTIONS

1. The Cutoff Criterion. The cutoff criterion must be established and unaltered throughout the analysis.
2. The Pre-Post Distribution. The pre-post distribution is a polynomial function.
3. Comparison Group Pretest Variance. Adequate numbers of pretest data points in the comparison group must be included to provide an appropriate estimation of the "true" relationship (i.e., pre-post regression line) for that group.
4. Continuous Pretest Distribution. A single continuous pretest distribution must be established for both groups and the division between groups determined by the cutoff.
5. Program Implementation. The program is uniformly delivered to all recipients in the treatment condition with the appropriate manipulation checks.

Pretest	Assignment	Group	Treatment	Posttest
O_A	C	1	X	O_2
O_A	C	2	-	O_2

Time ►

Note: O_A refers to the pre-assignment measure, and C refers to the cutoff score.

Design: Quasi-experimental research using a regression-discontinuity design

Null Hypothesis: For both groups, the linear relationship between the IV and DV is nonsignificant or equal to zero. In this example, $H_0: \beta_1 = 0; \beta_2 = 0$; or the slopes for Groups 1 and 2 are equal.

Statistic: ANCOVA and Linear Regression

Statistical Power: Researchers should use up to 2.75 times the number of participants than a 2-group pre- and posttest design with random assignment.

Specific Example

Research Question: To what extent do those who needed supplementary instruction and receive the remedial math program differ from those who did not go through the program on their state math achievement test?

Null Hypothesis: There is no difference in the scores on the state math achievement test (SMAT) for those who receive the remedial math program compared to those who did not go through the program.

Group	Pretest	Assignment	Treatment	Posttest
1 ($n = 48$)	SMAT	C	Remedial math program	SMAT
2 ($n = 44$)	SMAT	C	-	SMAT

DATA ENTRY

See data set for this particular example for an understanding on how to set up variables in Variable View and how to enter data into SPSS for the regression discontinuity design.

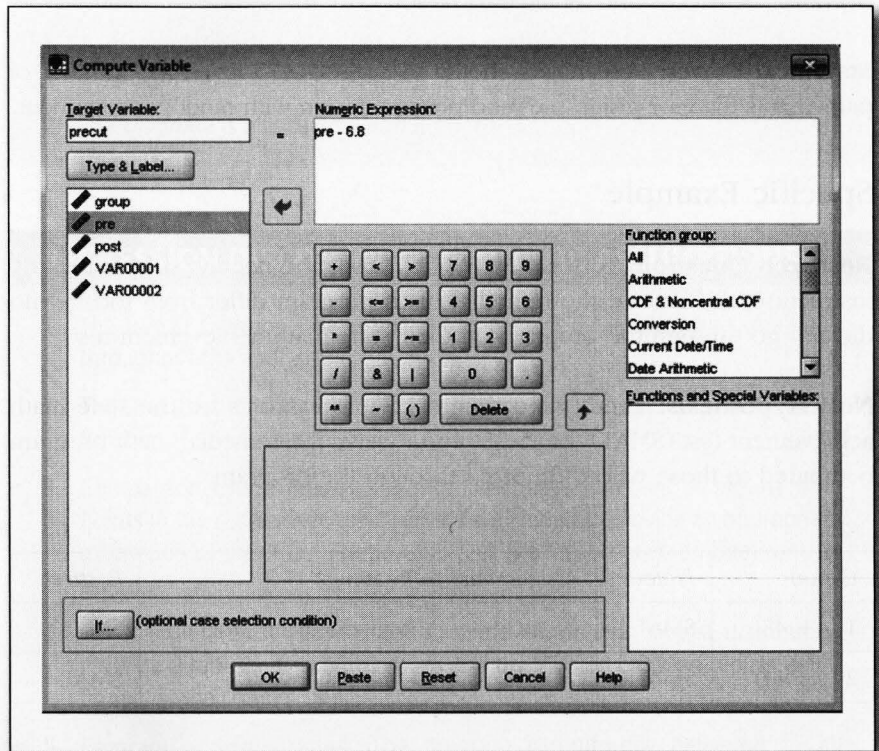
COMMAND PROMPTS FOR SPSS ♦

Data titled *RDdata.sav*

Transform the Data

Transform → Compute

Type “precut” in Target Variable and then move *pre* into Numeric Expression and then subtract 6.8. Click **OK**.



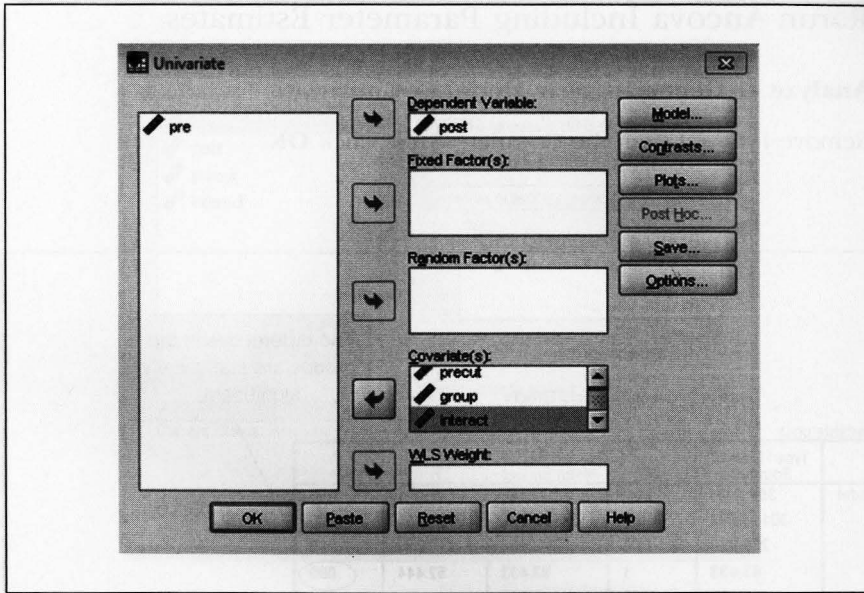
CREATING AN INTERACTION TERM

Transform → Compute

Similar to the first step, type “interact” in Target Variable and add $group * precut$ in the Numeric Expression box. Click **OK**.

ANCOVA INCLUDING PARAMETER ESTIMATES

Analyze → General Linear Model → Univariate



Click on **Model** and select **Custom**. Move *precut*, *group*, and *interact* into the Model box. Select Main Effect on the drop-down menu under Build Terms and then select Type I for the Sum of squares. Click **Continue**.

Click on **Options** and select Parameter Estimates. Click **Continue**. Click **OK**.

Tests of Between-Subjects Effects

Dependent Variable: post

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	355.021 ^a	3	118.340	73.929	.000
Intercept	30183.773	1	30183.773	18856.264	.000
precut	270.863	1	270.863	169.212	.000
group	83.433	1	83.433	52.122	.000
interact	.725	1	.725	.453	.503
Error	140.864	88	1.601		
Total	30679.658	92			
Corrected Total	495.885	91			

a. R Squared = .716 (Adjusted R Squared = .706)

The interaction is nonsignificant, indicating the two groups do *not* differ with respect to the slope of the regression line.

Rerun Ancova Including Parameter Estimates

Analyze → General Linear Model → Univariate

Remove *interact* from the Covariates box. Click **OK**.

Tests of Between-Subjects Effects

Dependent Variable: post

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	354.296 ^a	2	177.148	111.351	.000
Intercept	30183.773	1	30183.773	18972.854	.000
precut	270.863	1	270.863	170.259	.000
group	83.433	1	83.433	52.444	.000
Error	141.589	89	1.591		
Total	30679.658	92			
Corrected Total	495.885	91			

a. R Squared = .714 (Adjusted R Squared = .708)

The differences in the groups are statistically significant.

Parameter Estimates

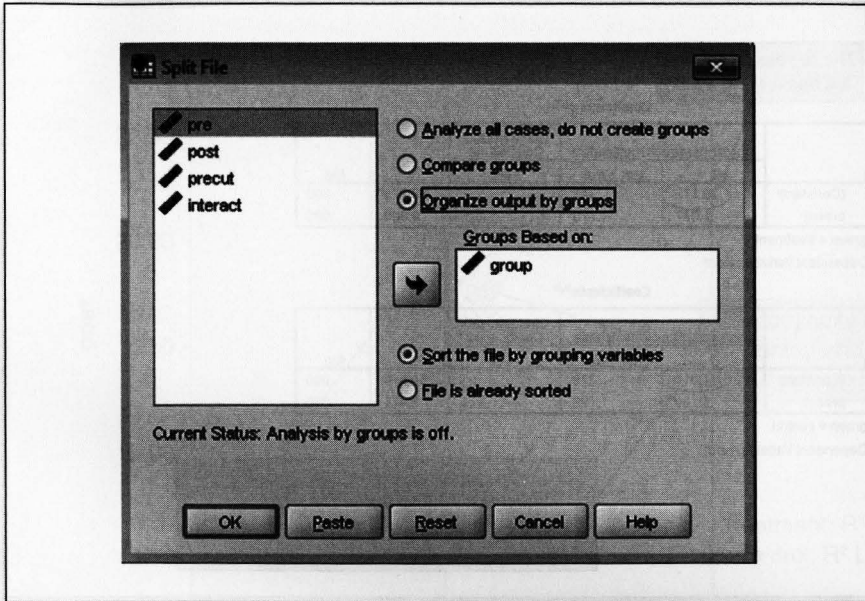
Dependent Variable: post

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	23.487	.702	33.439	.000	22.092	24.883
precut	1.124	.083	13.611	.000	.960	1.288
group	-3.185	.440	-7.242	.000	-4.059	-2.311

The slope for the group variable indicates that the change between the treatment and control is roughly 3.2 points on the SMAT.

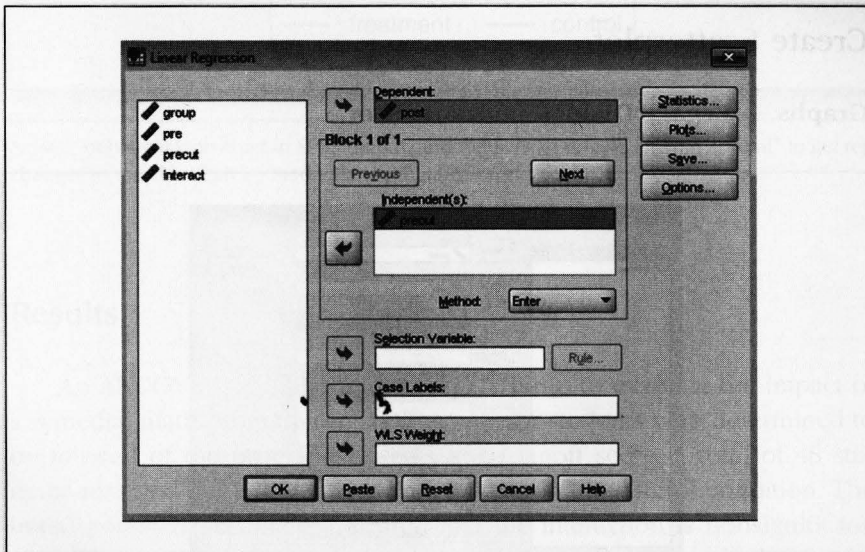
Separate the Regression Lines

Data → Split File



Linear Regression

Analyze → Regression → Linear



The regression line for the treatment is 20.2.

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	20.178	.343		58.778	.000
	precut	1.077	.110	.822	9.800	.000

a. group = treatment
b. Dependent Variable: post

Coefficients^{a,b}

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.010	.279		60.874	.000
	precut	1.190	.126	.824	9.411	.000

a. group = control
b. Dependent Variable: post

The regression line for the control group is 17. The treatment group is 20.2-17 = 3.2, as seen in the parameter estimates (i.e., estimated treatment effect).

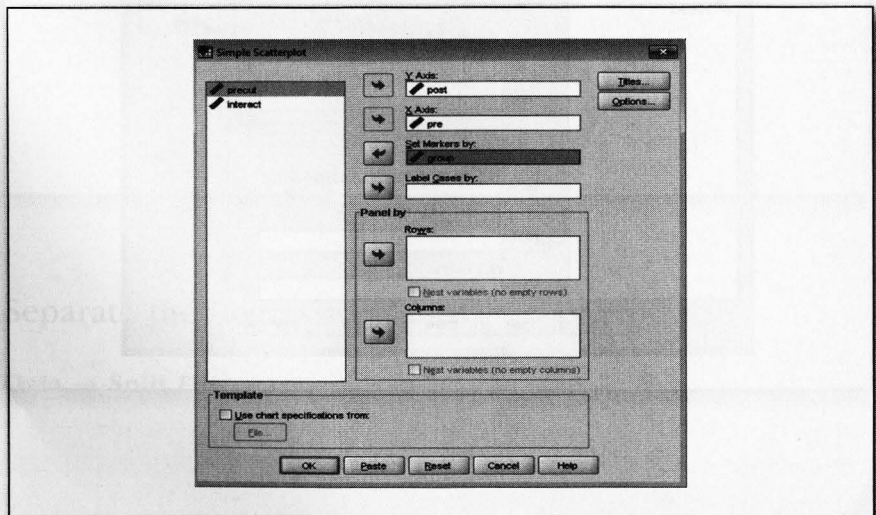
Separate the Regression Lines

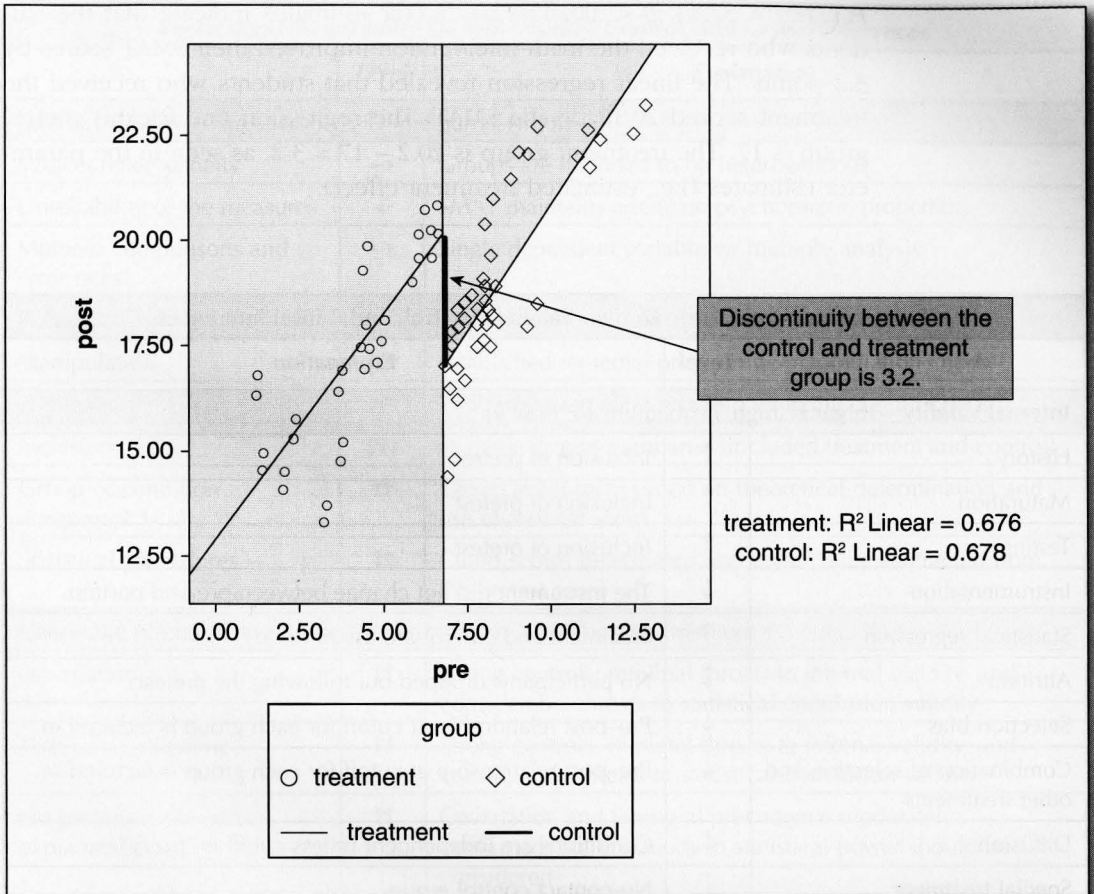
Data → Split File

Select *Analyze all cases, do not split groups*. Click **OK**.

Create Scatterplot

Graphs → Legacy Dialogs → Scatter/Dot





Note: Double click on chart in SPSS output, and then click on "Add Fit Line at Total" to get regression lines. Additional changes to the chart can be made using programs outside of SPSS.

Results

An ANCOVA and linear regression were run to examine the impact of a remedial math program on SMAT scores for students who determined to be in-need of the program. Based on the cutoff score, a total of 48 students received the intervention, and 44 were in the control condition. The initial parameter estimate indicated that the interaction is nonsignificant, $F(1, 88) = .453, p > .503$, indicating the two groups do *not* differ with respect to the slope of the regression line. The ANCOVA was rerun including the parameter estimates, indicating the differences between the intervention and control group were revealed to be statistically significant,

$F(1, 89) = 52.44, p < .00$. The parameter estimates indicate that the students who received the math intervention improved their SMAT scores by 3.2 points. The linear regression revealed that students who received the treatment scored 20.18 on the SMAT. The regression line for the control group is 17. The treatment group is $20.2 - 17 = 3.2$, as seen in the parameter estimates (i.e., estimated treatment effect).

<i>Regression-Discontinuity Design: Validity, Control, and Causal Inferences</i>		
	<i>Level</i>	<i>Explanation</i>
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
History	↑	Inclusion of pretest
Maturation	↑	Inclusion of pretest
Testing	↑	Inclusion of pretest
Instrumentation	↓	The instrument did not change between pre and posttest
Statistical regression	↑	Treatment and control are asymmetric
Attrition	↓	No participants dropped out following the pretest
Selection bias	↓	Pre–post relationship at cutoff for each group is factored in
Combination of selection and other treatments	↓	Pre–post relationship at cutoff for each group is factored in
Diffusion	↓	Conditions are independent unless cutoff is “fuzzy”
Special treatment	↓	No-contact control group
Sequencing effects	n/a	Between-subjects approach
Statistical Conclusion Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Low statistical power	↑	Requires roughly 2.75 times the participants than an equivalent 2-group pre–post design w/ random assignment
Assumption violation of statistical tests	↓	Cutoff criterion stable; the comparison group pretest variance and continuous pretest distribution confirmed
Error rate problem	↔	Combination of analysis
Restriction of range	↓	Treatment and control conditions
Extraneous variance in the experimental setting	↔	Strong levels of design control implemented
Inaccurate effect size estimation	↓	ANCOVA plus linear regression is the appropriate analysis
Variability in procedures	↓	Strong levels of design control implemented

Regression-Discontinuity Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Subject heterogeneity	↔	Groups are intended to be heterogeneous
Unreliability of the measures	↓	SMAT maintains adequate psychometric properties
Multiple comparisons and error rates	↔	Single dependent variable w/ multiple analysis
Control —Design and statistical control is (strong †††) (medium ††) (weak †)		
Manipulation	†††	Established remedial program w/ manipulation check
Elimination	†††	Established remedial program w/ manipulation check
Inclusion	†††	Multiple groups compared (included treatment and control)
Group or condition assignment	††	Group assignment based on theoretical determination and use of a cutoff score
Statistical procedures	††	Cutoff scores determined and cutoff value subtracted from the pretest score
Cause and Effect —Cause-effect determination is (strong †††) (medium ††) (weak †)		
Covariation	††	Strong controls, minimal threats to internal validity, and considerable threats to statistical conclusion validity
Temporal precedence	††	Strong controls, minimal threats to internal validity, and considerable threats to statistical conclusion validity
No plausible alternative explanation	††	Covariation and temporal precedence moderately confirmed, but violations in statistical power should be considered

Note: All threats are not created equal. Threats to construct and external validity should also be considered.

Example E8 *Ex Post Facto Design*

An industrial organizational consultant is interested in examining how a workshop designed to improve employee satisfaction in a factory setting differed between female and male employees. The intervention was already implemented by the company to all the employees, and certain design controls of interest were not used. Specifically, there were no control group or pretest measures; therefore, the consultant could not determine if employee satisfaction increased or changed as a result of the training. There was a discussion among employees that the training was biased toward the male employees. As a result, after the fact, the consultant was

able to collect posttest measures using the job satisfaction survey and then use gender as the matched-grouping criteria (statistical control) to portion out the data into two groups (males and females).

Group	Treatment		Posttest	Assignment	Group
1	X		O ₁	M _A	1
					2

Time ►

Note: M_A represents the post hoc matched-grouping criteria used as a means to include the desired participants in each condition. The assignment to conditions is conducted after the treatment has occurred. Depending on the research questions, matched-grouping criteria, and sample size, groups of three or more may result.

Design: Nonexperimental research using a between-subjects approach with an ex post facto design

Null Hypothesis: The adjusted population means for both groups are equal. In this example, H₀: $\mu_1 = \mu_2$ or the adjusted means of Groups 1 and 2 are equivalent.

Statistic: Independent Samples *t* Test

Statistical Power: Utilize G*Power Data Analysis

Specific Example

Research Question: How do males and females differ in terms of the scores on the job satisfaction survey following an employee satisfaction workshop?

Null Hypothesis: There is no difference in the scores on the Employee Satisfaction Survey between males and females.

Group	Treatment		Posttest	Assignment	Group
N = 54	Satisfaction training		Job satisfaction survey	Gender	Males (n = 24)
					Females (n = 30)

Data titled *ExPostFacto.sav*

Note: To review the scoring parameters for the job satisfaction survey, see Spector (1997).

Independent Samples *t* Test

Analyze → Compare Means → Independent Samples *t* Test

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Satisfaction	Male	24	148.0833	33.01504	6.73917
	Female	30	105.5000	23.84695	4.35382

Mean scores between males and females

The Levene statistic is significant, indicating homogeneity of variance is violated.

Independent Samples Test

Satisfaction	Equal variances assumed	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
Satisfaction	Equal variances assumed	4.728	52	.000	43.58333	7.74257	28.04114	59.11883
	Equal variances not assumed	5.432	40.997	.000	43.58333	8.62352	27.37522	59.79145

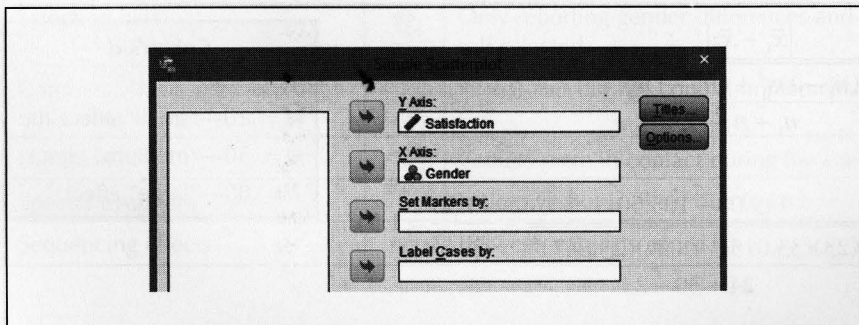
The *t* statistic shows significance ($p < .05$).

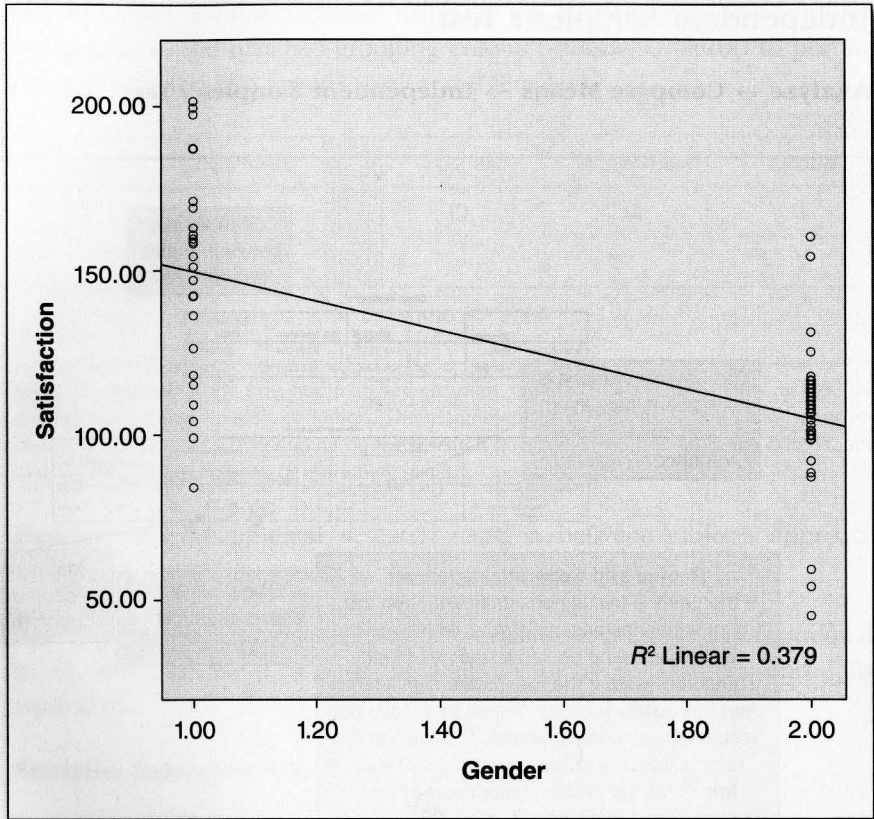
Males on average scored 44 points higher than females.

Pooled and Separate Variances
 If the Levene test is nonsignificant, then the "Equal variances assumed" row, which are the results from the pooled-variances *t* test, should be used. If the results are significant, then the results from the "Equal variances not assumed" row should be used. These are the results from the separate-variances *t* test, known as the Welch-Satterthwaite *t* test (see Zimmerman and Zumbo, 2009, for considerations with these tests).

Point Biserial Correlation Coefficient (Effect Size)

Graph → Legacy Dialogues → Scatter/Dot → Simple Scatter





Effect Size

SPSS does not provide a specific effect size calculation for the independent samples *t* test. The appropriate formula for the independent samples *t* test is the Cohen's *d*, which uses the pooled standard deviation and is also known as Hedges' *g*.

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

$$d = \frac{149.08 - 105.50}{\sqrt{\frac{(23)(33.015)^2 + (29)(23.847)^2}{24 + 30 - 2}}} = \frac{43.58}{28.27} = 1.54$$

Cohen's *d*

$d = .20$ —"small" effect
 $d = .50$ —"medium" effect
 $d = .80$ —"large" effect

The second option for an effect size calculation for the independent samples *t* test is the point biserial correlation.

Note: To run the r_{pb} , select **Analyze** → **Correlate** → **Bivariate**, and in the Variables box, enter *Gender* and *Satisfaction*.

Results

An independent samples *t* test was used to compare the scores on the job satisfaction survey (JSS) between males and females following an employee satisfaction workshop. The Levene statistic was significant, indicating the homogeneity of variance assumption was violated. Therefore, the results from the separate-variances *t* test (equal variances *not* assumed) was used. The differences between the two conditions were found to be statistically significant $t(40.59) = 5.432, p < .000$. These findings indicate that the differences between males ($M = 149.08, SD = 33.01$) and females ($M = 105.5, SD = 23.84$) did not happen by chance. The *Mean* difference between the two groups was about 44 points. Practically speaking, the effect size calculation revealed a “large” effect ($d = 1.54$). The point biserial correlation also revealed a “large” effect ($r_{pb} = -.615; R^2 = -.379$).

Ex Post Facto Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
History	↑	Inclusion of pretest
Maturation	↑	Inclusion of pretest
Testing	↑	Inclusion of pretest
Instrumentation	↓	The instrument did not change between pre- and posttest
Statistical regression	↑	Cannot determine based on data
Attrition	↑	No participants dropped out following the pretest
Selection bias	↔	Only reporting gender differences and groups were not self-selected
Combination of selection and other treatments	↑	Researcher did not control implementation of program
Diffusion	↑	Groups were in contact during the implementation phase
Special treatment	↑	Unknown but is likely due to lack of design controls
Sequencing effects	n/a	Between-subjects approach

(Continued)

(Continued)

Ex Post Facto Design: Validity, Control, and Causal Inferences		
	Level	Explanation
Statistical Conclusion Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Low statistical power	↑	Number of participants in each group is low
Assumption violation of statistical tests	↔	Basic assumptions were met
Error rate problem	↔	Basic analysis
Restriction of range	↓	Two groups divided by gender
Extraneous variance in the experimental setting	↑	No design controls implemented
Inaccurate effect size estimation	↔	Basic analysis
Variability in procedures	↑	No design controls implemented
Subject heterogeneity	↔	Groups are heterogeneous
Unreliability of the measures	↓	JSS is a valid and reliable instrument
Multiple comparisons and error rates	↔	Basic analysis
Control —Design and statistical control is (strong †††) (medium ††) (weak †)		
Manipulation	†	Not implemented
Elimination	†	Not implemented
Inclusion	†	Not implemented
Group or condition assignment	†	Implemented after the fact
Statistical procedures	††	Groups determined by gender classification
Cause and Effect —Cause-effect determination is (strong †††) (medium ††) (weak †)		
Covariation	†	Weak controls, high threats to internal validity, and considerable threats to statistical conclusion validity
Temporal precedence	†	Weak controls, high threats to internal validity, and considerable threats to statistical conclusion validity
No plausible alternative explanation	†	Cannot determine based on current design

Note: Ex post facto designs are nonexperimental research, but aspects of internal validity do apply, as the application is intended to make quasi-experimental research from a nonexperimental application. Therefore, results should be interpreted with caution, as noted by demonstrating the weaknesses in determining causation.

Appendix F

Qualitative Data Analysis

In this appendix, we provide clear guidelines, demonstrating how to organize, prepare, code, and present qualitative data. This appendix includes four sections: The first addresses the preparation of qualitative data; the second presents the use of memoing; the third covers the functions of codes and coding used throughout the qualitative research process; and the fourth focuses on presenting the data.

Although there are different steps in the process of analyzing data, depending on the specific qualitative approach, there are some common core analytic elements. The core steps include reducing the data into meaningful segments and assigning names for the segments, putting the codes into broader categories (themes), and displaying (comparing and contrasting) the data with various visual aids (tables and charts). Thus, always keep in mind these four steps:

1. First, *prepare* and organize the data for analysis.
2. Then, use *memos* as a form of preliminary analysis (initial process of exploring the database).
3. Next, reduce the data into themes through a process of *coding* and condensing the codes.
4. Finally, *present* the data in figures, tables, and narrative.

♦ 1. PREPARATION

Before beginning analysis, it has to be in a format suitable for coding. The most common form of collecting data in qualitative research is conducting interviews, which are then transcribed into text-based data.

It is not always required to transcribe an interview or the field notes to carry out analysis. Some forms of analysis can be done directly from the recording or notes. This allows for a broad-based or general exploration of the data, rather than a more direct focus on all of the specific details of what was said. Keep in mind that most qualitative methodologies require the analysis of the specific details of exactly what and how things were said.

When the design dictates that it is not necessary to analyze everything that the participant said or everything that was said in perfect detail, researchers often transcribe parts of the interview and write notes on the rest. This is quicker and cheaper than a full transcription. It also allows the researcher to focus on the bigger picture. The downside to this approach is that the transcribed parts can be out taken out of context, and it makes it hard to interpret without constant reference to the tape or the video. Also, this approach adds more subjectivity to the process—in other words, what is thought as significant at the time of transcribing may take on a different meaning during analysis.

If a survey was used or notes were taken (rather than an interview), then the data may be organized into computer files. Put the data into the appropriate text units—this may be words, sentences, paragraphs, or an entire story. If a narrative design was used, the story should be put into chronological order, then “retold” with the participant to ensure the trustworthiness of the data.

If audio recording was used, the exhaustive process of transcribing is usually required. The transcription process may also include the process of note-taking. Accurate transcription should indicate any nonverbal communication (e.g., tone, mannerisms, emotional content, or contextual factors). Make sure to take notes on how feelings and meanings were expressed nonverbally during the interview. For instance, perhaps the participant stated, “Sure, everybody wants kids.” However, this might be said in a variety of ways, conveying different meanings other than what the words imply.

“Sure, everybody wants kids.” (This was stated with conviction; the participant really believes this.)

“Sure, everybody wants kids.” (This was stated sarcastically, while the participant laughed and blushed; clearly the participant does not believe this.)

“Sure, everybody wants kids?” (This was asked as a question; the participant is really not sure if he believes this or not.)

This process takes careful active listening to fully understand the meaning behind the words that are recorded. Jotting down memos or field notes during the data collection process also is critical to make sure nonverbal communication is accurately interpreted.

Example of Two Types of Transcription

Semi Transcription

I have to go work all day, which makes it difficult to spend time with my kids. My wife is always complaining that I am not around enough.

Full Transcription

I work from 7 in the morning until 9 at night. By the time I get home, my kids are asleep, and my wife is worn out. My wife has to raise our three kids pretty much by herself during the week. I make sure to spend time with my kids on the weekend, but I get little time to be alone with my wife. She complains all week that I don't spend enough time with her, but it is hard to find time.

2. MEMOING ♦

Memoing is the act of recording reflective notes about what you (the researcher) are learning from the data:

- Use memos to record your ideas about concepts and their relationships within your data.
- These memos (notes) allow you to begin to work out some hypothesis regarding a category and relationships between categories.
- Memos increase the *credibility* of your findings.

Memoing can be used throughout the process of collecting (similar to field notes) and analyzing (analytic memoing) qualitative data. The use of memoing during data collection is useful to help jog the researcher's memory of important concepts and connections between concepts or ideas

in the data. Later, analytic memoing is conceptually similar to the preliminary data analyses conducted in quantitative methods.

In fact, Guba (1981) developed four criteria for qualitative research that parallel constructs associated with quantitative inquiry:

1. Credibility (internal validity)
2. Transferability (external validity/generalizability)
3. Dependability (reliability)
4. Confirmability (objectivity)

Credibility is often operationalized as confidence in the data and the analysis. Credibility, as Guba (1981) noted, is synonymous with internal validity in quantitative research. One of the means to enhance the credibility is to take notes (memos). These notes strengthen the defensibility of the results.

As mentioned, an important use of memoing is for preliminary qualitative data analysis purposes. In many qualitative approaches, data analysis happens at two levels: textual and conceptual, which are seldom clearly delineated. The textual level is the preliminary stage of analyses using memoing techniques, while the conceptual level involves the primary analysis of the qualitative data, using various coding procedures (however, memoing is often used during this process as well).

The textual level relates to the reading of all of the field data and memoing throughout. Qualitative researchers emphasize that memoing is prioritized—in other words, when an idea occurs the researcher pauses and records it. As the researcher forms ideas, they are written down as memos. The literature is replete with different types of memoing. Theoretical memoing is about attempts to derive meaning from the data. Methodological or operational memoing comprises reminders, instructions, or critiques that the researcher writes himself as the analysis unfolds. However, many qualitative researchers choose not to bind themselves by these artificial parameters and just memo, perhaps organizing them using specific taxonomy later on in the process using these qualifiers.

The conceptual level of memoing involves theorizing about concepts and themes and the relationships between these. Here, the researcher begins to fit together the different pieces of the puzzle. Some pieces may not fit at first, but upon reexamination and contemplation, they may find a place within the overall story being told.

The analysis is mostly simultaneous or parallel and often entails recurring phases of data collection, coding, memoing, and sorting. Memos help

the researcher to see the “forest through the trees” by keeping the researcher focused on what the data is conveying conceptually.

- Memoing involves total creative freedom.
- There are no rules regarding writing, grammar, or style.
- A memo is purely an instrument to capture the outflow of ideas, insights, and observations.
- When the researcher writes thoughts down, they become concrete, and they are recorded.
- There are no wrong or poorly written memos.
- Each researcher develops his or her own style.
- Memos evolve and increase in complexity, density, clarity, and accuracy as the data analysis progresses.
- Memos written later may negate, amend, extend, and/or clarify earlier written ones.
- Memos keep the researcher embedded in the empirical reality and contribute to the trustworthiness of qualitative research. Trustworthiness is synonymous with reliability in quantitative research.

Diagrams are graphic memos and may play a very important conceptual role. Diagrams are visual devices that depict something. They illustrate the density and complexity of the qualitative analysis. A diagram may allow for the discovery of gaps and flaws in the relationships of categories. The use of listings early on in the process helps in the later construction of visual diagrams.

Thus, the act of recording reflective notes during data collection and analysis is called *memoing*. These memos add to the credibility and trustworthiness of qualitative research. Memoing aids the analysis in that the researcher records the meanings derived from the data. Although there are no strict rules regarding memoing, the memos should (a) contain one idea, (b) be dated, and (c) be referenced. Remember that memos evolve as the research proceeds and help the researcher conceptualize, organize, and code the data.

3. CODING ♦

Coding is the process of exploring the data for themes, ideas, and categories and then marking similar passages of text with a code label so that they can easily be retrieved at a later stage for further comparison and analysis.

Coding the data makes it easier to search the data, to make comparisons, and to identify any patterns that require further investigation. Simply put, a code is often a word (or short phrase) that assigns a summative or salient attribute to some portion on the data.

Codes can be based on (a) themes or topics, (b) concepts, (c) terms or phrases, and (d) keywords found in the data. Usually, passages of text are coded, but they can also be sections of an audio or video recording or parts of images. Any portions of the data that are coded the same way should demonstrate the same theme, concept, or idea.

The codes are given meaningful names that provide an indication of the idea or concept that reinforces the theme or category. Any parts of the data that relate to a code topic are coded with the appropriate label. This process of coding (associating labels with the text and images) involves close reading of the text (or close inspection of the video or images). If a theme is identified from the data that does not quite fit the codes already existing, then a new code is created. The number of codes will evolve and grow as more topics or themes become apparent.

Open Coding

At this first level of coding, the focus is on distinct concepts and categories in the data, which will form the basic units of the analysis. The data is organized into first-level concepts (master headings) and second-level categories (subheadings).

The use of highlights to distinguish concepts and categories is often helpful. For example, if a participant continually refers to bullying, each time bullying or something related to bullying is mentioned, it could be highlighted with the same color. Bullying would become a concept, and other things related to bullying would become categories highlighted with the same color. Each broad construct or category should be highlighted with a different color. Once this stage is complete, the transcripts will have multiple colors of highlighted text. These may be organized as an outline, with concepts being main headings and categories labeled as subheadings within these main headings.

Axial Coding

The next step is axial coding. Here, concepts and categories (developed during open coding) are used while rereading the text to (a) confirm that concepts and categories accurately represent interview responses and (b) explore how concepts and categories are related. This is accomplished by asking questions such as the following:

1. What conditions cause or influence categories?
2. What is the context?
3. What are the associated effects or consequences?

For instance, if bullying is a concept that emerged with two categories (cyberbullying and mobbing), an axial code might be a phrase such as “Cyberbullying is occurring within the workplace.” In essence, the code delineates the context of the concept and categories, suggesting the need for a new category (cyberbullying in organizations). Thus, axial coding is a more focused approach of looking at the data, ensuring that the researcher has identified all important aspects.

Create a Table

Final concepts and categories may be moved into a data table. The table may list the major concepts and categories. The table may be followed by a narrative delineating the categories.

The concepts, categories, and subcategories may progress toward the thematic, conceptual, and theoretical.

4. PRESENTING ♦

When preparing to report, the first thing to consider is how to organize the findings. The findings represent the story that answers the central research question and subquestions guiding the study. Thus, organize the findings in a way that makes logical sense and keeps the focus on those guiding questions. Findings are usually organized by research question or by theme.

Findings should provide sufficient evidence from the data to support the conclusions. Evidence includes, but is not limited to, quotations from interviews and excerpts from observations and documents.

Ethically, it is important to have confidence in the findings and account for counterevidence (evidence that contradicts your primary finding). If there is not enough evidence to back up a primary thought or idea, it should not be included.

Findings should be linked back to the theoretical lens or conceptual framework.

Numbers, in the form of descriptive statistics, may help to convey the prevalence of a finding. Although this is qualitative inquiry, do not shy

away from the use of quantitative data if it helps provide context and support to the findings.

The following ideas are useful to consider when presenting qualitative findings:

1. It is not necessary to present every minute detail; instead, discuss the main themes as they relate to the central and subquestions.
2. Triangulate the data; compare and contrast findings by individual or group.
3. Use brief quotes to illustrate a particular point.

When presenting qualitative findings, remember that the data are (a) subjective, (b) interpretative, (c) descriptive, (d) holistic, and (e) copious. A common way to present findings is to structure the arrangement of the data around the categories or themes that have emerged. The themes or categories may be presented as sections with relevant subsections. Use quotes to delineate or support findings. Think about the possibility of using some quantitative data in the form of descriptive statistics to provide context to your qualitative findings.

It is a good idea to begin a presentation of qualitative findings with a brief explanation of how data were processed and coded, as well as how data exemplars were chosen.

In addition, present the different levels of coding and how data were categorized and subcategorized to help the reader visualize the structure of data coding. It may be useful to quantify the initial coding results to demonstrate the relative importance of different constructs found in the data.

Finally, it is important to include quotes that reflect the concept being coded. This also provides a valid connection between the open-text comments analyzed and the coding structure used in the analysis.

Chenail (1995) discussed the importance of juxtaposition when presenting qualitative findings:

An important concept to keep in mind as you juxtapose your data and your talk about the data is that of rhythm. By rhythm, I mean for you to create a template for re-presenting your data so that there is a recognizable pattern throughout the Analysis or Findings section of your paper. In this way, the readers can begin to read in a rhythm.

To accomplish this rhythm, you need to structure each phase of your data re-presentation in a similar pattern. For example, the following is a common way in which your findings can be displayed:

*Section Heading**Present the Distinction or Finding**Introduce the First Data Exemplar of this Distinction**Display the First Data Exemplar of this Distinction**Comment Further on the First Data Exemplar of this Distinction**Make Transition to Second Data Exemplar of this Distinction**Display the Second Data Exemplar of this Distinction**Comment Further on the Second Data Exemplar of this Distinction**Make Transition to the Next Data Exemplar of this Distinction and**Repeat the Pattern Until the Closing of this Section*

As you write (and re-write) your Findings or Analysis section, having a pattern to structure your text will greatly improve your ability to lay out the data upon which you want to comment and then for you to weave your comments throughout the narrative.

As a matter of fact, that is exactly how I construct my Findings sections. After I have selected my data exemplars (e.g., quotes from a transcript), I arrange them out in the word processing file (see the next section on Data Presentation Strategies for different ideas as to how to do this step). Then, I go from exemplar to exemplar adding my comments. Sometime I call these steps the “Tarzan Process,” because I think of the quotes as vines in the jungle. As I maneuver myself from one quote to the next, I imagine myself as Tarzan swinging from one vine to another. It’s a great way to travel and a fun way to conceptualize the data re-presentation process.

Just as the patterning was helpful for you as you wrote up your findings, it will also serve your readers well as they begin to navigate through your paper. The soon-to-be-familiar rhythm of your presentation style will serve as an involvement strategy as the readers will grow accustomed to your pace of data re-presentation. In addition, it will make it easier for them to go from section to section. Although the data will be changing, the pattern will remain the same. In this manner, cross-section comparisons can be made more readily by the readers, which, in my opinion, helps to make the whole paper reading process much more coherent.

By keeping to this or another similarly rhythmic pattern, you can help to bring some simplicity to the complexity of data

re-presentation. Throughout all the steps entailed in conducting a qualitative research study, you must always attempt to build in some sort of simplicity. Without it, both you and the reader will be overcome and you all will end up drowning in a sea of endless data. (sec. 3, para. 5–10)

Chenali (1995) goes on to offer data presentation strategies:

The following are a few examples of the myriad ways that qualitative data can be presented:

Natural—The data are presented in a shape that resembles the phenomenon being studied. For instance, if the data are excerpts from a therapy session, present them in a sequential order or in an order that re-presents the flow of the session itself.

Most Simple to Most Complex—For the sake of understanding, start the presentation of data with the simplest example you have found. As the complexity of each example or exemplar presented increases, the reader will have a better chance of following the presentation. Erving Goffman's work is a good example of this style.

First Discovered/Constructed to Last Discovered/Constructed—The data are presented in a chronicle-like fashion, showing the course of the researcher's personal journey in the study. This style is reminiscent of an archeological style of presentation: What was the first "relic" excavated, then the second, and so forth.

Quantitative-Informed—In this scheme, data are presented according to strategies commonly found in quantitative or statistical studies. Data are arranged along lines of central tendencies and ranges, clusters, and frequencies.

Theory-Guided—Data arrangement is governed by the researcher's theory or theories regarding the phenomenon being re-presented in the study. For instance, a Marxist-informed researcher might present data from a doctor–patient interview in terms of talk that shows who controls the means for producing information in the interaction, talk that illustrates who is being marginalized, and so forth. In clinical qualitative research, this approach is quite prevalent as clinicians organize the data in terms of their understandings of how doctor–patient, or nurse–patient, and therapist–client interact.

Narrative Logic—Data are arranged with an eye for storytelling. Researchers plot out the data in a fashion that allows them to transition from one exemplar to another, just as narrators arrange details in order to best relate the particulars of the story.

Most Important to Least Important or From Major to Minor—Like the journalistic style of the inverted pyramid, the most important “findings” are presented first, and the minor “discoveries” come last.

Dramatic Presentation—This one is the opposite of the inverted pyramid style. With the dramatic arrangement scheme, researchers order their data presentation so as to save the surprises and unforeseen discoveries for last.

No Particular Order—As it sounds, data are arranged with no particular conscious pattern in mind, or the researcher fails to explain how or why the data are displayed the way they are. (sec. 4, para. 3)

Appendix G

Validity, Control, and Causal Inferences Checklist

Validity, Control, and Causal Inferences		
	<i>Level</i>	<i>Explanation</i>
Internal Validity —Threat is (high ↑) (medium ↔) (low ↓)		
History		
Maturation		
Testing		
Instrumentation		
Statistical regression		
Attrition		
Selection bias		
Combination of selection and other treatments		
Diffusion		
Special treatment		
Sequencing effects		

Statistical Conclusion Validity —Threat is (high ↑) (medium ↔) (low ↓)		
Low statistical power		
Assumption violation of statistical tests		
Error rate problem		
Restriction of range		
Extraneous variance in the experimental setting		
Inaccurate effect size estimation		
Variability in procedures		
Subject heterogeneity		
Unreliability of the measures		
Multiple comparisons and error rates		
Control —Design and statistical control is (strong III) (medium II) (weak I)		
Manipulation		
Elimination		
Inclusion		
Group or condition assignment		
Statistical procedures		
Cause and Effect —Cause-effect determination is (strong III) (medium II) (weak I)		
Covariation		
Temporal precedence		
No plausible alternative explanations		

Note: Not all threats to validity are created equal. Construct validity should also be considered along with external validity, depending on the objectives of the researchers. The checklist can be considered a subjective-objective approach to reviewing the overall quality of the methodological application of an examination.

Appendix H

CONSORT 2010: Checklist and Guidelines for Reporting Parallel Randomised Trials

<i>Section/Topic</i>	<i>Item No.</i>	<i>Checklist Item</i>	<i>Reported on Page No.</i>
Title and abstract			
	1a	Identification as a randomised trial in the title	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	
	2b	Specific objectives or hypotheses	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial), including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	

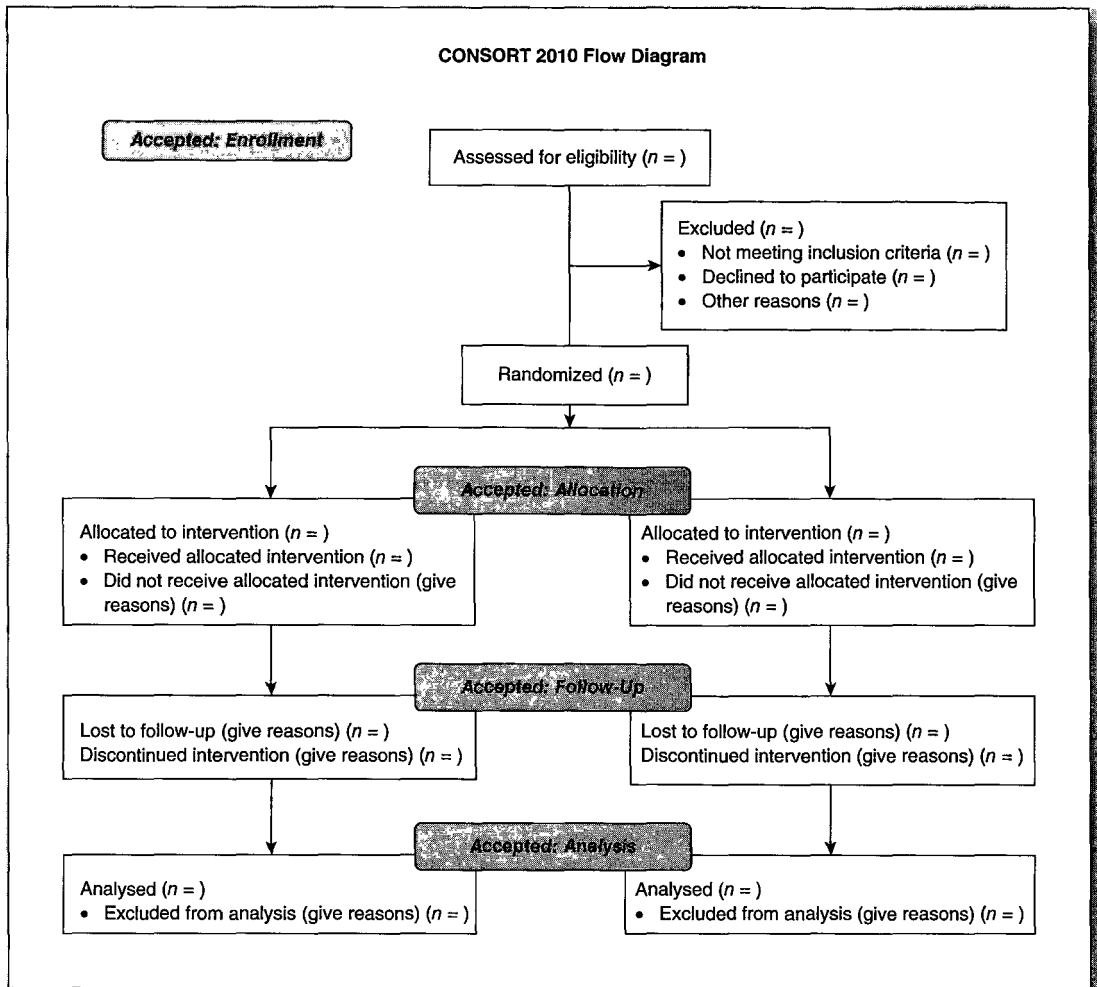
<i>Section/Topic</i>	<i>Item No.</i>	<i>Checklist Item</i>	<i>Reported on Page No.</i>
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation Concealment Mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomisation, together with reasons	

(Continued)

(Continued)

Section/Topic	Item No.	Checklist Item	Reported on Page No.
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance, see CONSORT for harms)	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalizability	21	Generalizability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

Note: We strongly recommend reading this statement in conjunction with the *CONSORT 2010 Explanation and Elaboration* for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomized trials, noninferiority and equivalence trials, nonpharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming; for those and for up-to-date references relevant to this checklist, see www.consort-statement.org.



Source: Schultz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Research Methods & Reporting*, 340, 698–702.

Note: MS-Word versions of the checklist and flowchart can be found at <http://www.equator-network.org/reporting-guidelines/consort/>.

References

- Abrams, L. S., Shannon, S. K. S., & Sangalang, C. (2008). Transition services for incarcerated youth: A mixed methods evaluation study. *Children and Youth Services Review, 30*, 522–535.
- Acee, T. W., & Weinstein, C. E. (2010). Effects of value-reappraisal intervention on statistics students' motivation and performance. *Journal of Experimental Education, 78*(4), 487–512.
- Acosta, S., Goltz, H. H., & Goodson, P. (2015). Autoethnography in action research for health education practitioners. *Action Research, 13*(4), 411–431. doi:10.1177/1476750315573589.
- American Psychological Association. (2010). *Concise rules of APA style* (6th ed.). Washington, DC: Author.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Anderson, R. B., Doherty, M. E., & Friedrich, J. C. (2008). Sample size and correlational inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 929–944.
- Baker, M. (2015, August 27). Over half of psychology studies fail reproducibility test. *Nature*. doi:10.1038/nature.2015.18248.
- Basadur, M., Graen, G. B., & Scandura, T. A. (1986). Training effects on attitudes toward divergent thinking among manufacturing engineers. *Journal of Applied Psychology, 71*(4), 612–617.
- Baylor, A. L., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior, 25*, 450–457.
- Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo* (2nd ed.). Thousand Oaks, CA: Sage.
- Beeke, S., Wilkinson, R., & Maxin, J. (2007). Individual variation in agrammatism: A single case study of the influence of interaction. *Journal of Language and Communication Disorders, 42*(6), 629–647.
- Berman, H., Mulcachy, G. A., Forchuk, C., Edmunds, K. A., Haldenby, A., & Lopez, R. (2009). Uprooted and displaced: A critical narrative study of homeless, aboriginal, and newcomer girls in Canada. *Issues in Mental Health Nursing, 30*(7), 418–430.
- Bernard, R. S., Cohen, L. L., & Moffett, K. (2008). A token economy for exercise adherence in pediatric cystic fibrosis: A single-subject analysis. *Journal of Pediatric Psychology, 34*(4), 354–365.

- Birks, M., Chapman, Y., & Francis, K. (2008). Memoing in qualitative research: Probing data and processes. *Journal of Research in Nursing, 13*(1), 68–75.
- Bohannon, J. (2015). Many psychology papers fail replication test. *Science, 349*(6251), 910–911.
- Bortree, D. S., & Seltzer, T. (2009). Dialogic strategies and outcomes: An analysis of environmental advocacy groups' Facebook profiles. *Public Relations Review, 35*(3), 317–310.
- Box, G. E. P. (2009). *Statistical control by monitoring and adjustment, statistics for experimenters: Design, innovation, and discovery—Wiley Series in Probability and Statistics* (2nd ed.). Hoboken, NJ: Wiley.
- Bryant, D. P., Bryant, B. R., Gersten, R., Scammacca, N., & Chavez, M. M. (2008). Mathematic intervention for first- and second-grade students with mathematics difficulties: The effects of tier 2 intervention delivered at booster lessons. *Remedial and Special Education, 29*(1), 20–31.
- Burgess, G., Grogan, S., & Burwitz, L. (2006). Effects of a 6-week aerobic dance intervention on body image and physical self-perceptions in adolescent girls. *Body Image, 3*, 57–66.
- Butler-Kisber, L. (2010). *Qualitative inquiry: Thematic, narrative and art-informed perspectives*. Thousand Oaks, CA: Sage.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54*, 297–312.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Skokie, IL: Rand McNally.
- Cernin, P. A., & Lichtenberg, P. A. (2009). Behavioral treatment for depressed mood: A pleasant events intervention for seniors residing in assisted living. *Clinical Gerontologist, 32*, 324–331.
- Chakraborti, C., Boonyasai, R. T., Wright, S. M., & Kern, D. E. (2008). A systematic review of teamwork training interventions in medical student and resident education. *Journal of General Internal Medicine, 23*(6), 846–853.
- Chao, P., Bryan, T., Burstein, K., & Ergul, C. (2006). Family-centered intervention for young children at-risk for language and behavior problems. *Early Childhood Education Journal, 34*(2), 147–153.
- Chapin, M. H., & Holbert, D. (2009). Differences in affect, life satisfaction, and depression between successfully and unsuccessfully rehabilitated persons with spinal cord injuries. *Rehabilitation Counseling Bulletin, 53*(1), 6–15.
- Charmaz, K. (2014). *Constructing grounded theory: A practical guide through qualitative analysis* (2nd ed.). London, England: Sage.
- Chenail, R. (1995, December). Presenting qualitative data. *The Qualitative Report, 2*(3). Retrieved from <http://www.nova.edu/ssss/QR/QR2-3/presenting.html>
- Chenail, R. (2010). Getting specific about qualitative research generalizability. *Journal of Ethnographic & Qualitative Research, 5*, 1–11.
- Chevalier, J. M., & Buckles, D. J. (2013). *Participatory action research: Theory and methods for engaged inquiry*. London, England: Routledge.

- Clandinin, D. J., & Connelly, F. M. (2004). *Narrative inquiry: Experience and story in qualitative research*. San Francisco, CA: Wiley.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Collins, K. M. T., & O’Cathain, A. (2009). Ten points about mixed methods research to be considered by the novice researcher. *International Journal of Multiple Research Approaches*, 3(1), 2–7.
- Comaskey, E. M., Savage, R. S., & Abrami, P. (2009). A randomized efficacy study of web-based synthetic and analytic programmes among disadvantaged urban kindergarten children. *Journal of Research in Reading*, 32(1), 92–108.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.
- Cook, T. D., & Steiner, P. M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pretest measures of outcome, of unreliable measurement, and of mode of data analysis. *Psychological Methods*, 15(1), 56–68.
- Coppola, L., Mastrolorenzo, L., Coppola, A., De Biase, M., Adamo, G., Forte, R., . . . Riccardi, A. (2013). QT dispersion in mild cognitive impairment: A possible tool for predicting the risk of progression to dementia? *International Journal of Geriatric Psychiatry*, 28(6), 632–639.
- Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2012). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.
- Creswell, J. W. (2014). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (5th ed.). Upper Saddle River, NJ: Pearson.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.
- Cribbie, R. A., Arpin-Cribbie, C. A., & Gruman, J. A. (2010). Tests of equivalence for one-way independent groups designs. *Journal of Experimental Education*, 78, 1–13.
- Dasgupta, T., Pillai, N. S., & Rubin, D. B. (2014). Causal inference from 2^k factorial designs using potential outcomes. *Journal of the Statistical Society: Series B (Statistical Methodology)*, 77(4), 727–753.
- Dennis, J. K. (2003). Problem-based learning in online vs. face-to-face environments. *Education for Health*, 16(2), 198–209.
- DeVellis, R. F. (2011). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage.
- Dixon, M. R., Jackson, J. W., Small, S. L., Horner-King, M. J., Lik, N. M. K, Garcia, Y., & Rosales, R. (2009). Creating single-subject design graphs in Microsoft Excel™ 2007. *Journal of Applied Behavior Analysis*, 42(2), 277–293.

- Drechsler, J. (2015). Multiple imputation of multilevel missing data-rigor versus simplicity. *Journal of Educational and Behavioral Statistics, 40*(1), 69–95.
- Elliott, M. S., & Williams, D. I. (2002). A qualitative evaluation of an employee counselling service from the perspective of client, counselor, and organization. *Counselling Psychology Quarterly, 15*(2), 201–208.
- Erdogan, Y., Aydin, E., & Kabaca, T. (2008). Exploring the psychological predictors of programming achievement. *Journal of Instructional Psychology, 35*(3), 264–270.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioral Research Methods, 39*(2), 175–191.
- Fetterman, D. M. (2009). *Ethnography: Step-by-step* (3rd ed.). Thousand Oaks, CA: Sage.
- Fincham, J. E. (2008). Response rates and responsiveness for surveys, standards, and the journal. *American Journal of Pharmaceutical Education, 72*(2), 1–3.
- Fink, A. G. (2012). *How to conduct surveys: A step-by-step guide* (5th ed.). Thousand Oaks, CA: Sage.
- Fletcher, A., & Marchildon, G. (2014). Using the Delphi method for qualitative, participatory action research in health leadership. *International Journal of Qualitative Methods, 13*, 1–18.
- Forchuk, C., Norman, R., Malla, A., Martin, M., McLean, T., & Cheng, S., . . . Gibney, C. (2002). Schizophrenia and the motivation for smoking. *Perspectives in Psychiatric Care, 38*(2), 41–49.
- Fowler, F. J. (2013). *Survey research methods: Applied social research methods* (5th ed.). Thousand Oaks, CA: Sage.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Boston, MA: Pearson.
- Ganz, J. B., & Flores, M. M. (2009). The effectiveness of direct instruction for teaching language to children with autism spectrum disorders: Identifying material. *Journal of Autism and Developmental Disorders, 39*, 75–83.
- Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). London, England: Routledge.
- Geertz, C. (1998). The world in pieces: Cultural politics at the end of the century. *FOCAAL: European Journal of Anthropology, 32*, 91–117.
- Gier, V. S., Kreiner, D. S., & Natz-Gonzalez, A. (2009). Harmful effects of preexisting inappropriate highlighting on reading comprehension and metacognitive accuracy. *Journal of General Psychology, 136*(3), 287–300.
- Giuffre, M. (1997). Designing research: Ex post facto designs. *Journal of PeriAnesthesia Nursing, 12*(3), 191–195.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Hawthorne, NY: Aldine Transaction.
- Glass, V. G., Wilson, V. L., & Gottman, J. M. (2008). *Design and analysis of time-series experiments*. Charlotte, NC: Information Age.

- Glombiewski, J. A., Sawyer, A. T., Gutermann, J., Koenig, K., Rief, W., & Hofmann, S. G. (2010). Psychological treatments for fibromyalgia: A meta-analysis. *Pain, 151*(2), 280–295.
- Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597–607.
- Green, M., & Cifuentes, L. (2008). An exploration of online environments supporting follow-up to face-to-face professional development. *Journal of Technology and Teacher Education, 16*(3), 283–306.
- Green, S., & Salkind, N. J. (2013). *Using SPSS for Windows and Macintosh* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Groenwold, R. H. H., Donders, A. R. T., Roes, K. C. B., Harrell, F. E., & Moons, K. G. M. (2012). Dealing with missing outcome data in randomized trials and observational studies. *American Journal of Epidemiology, 175*(3), 210–217.
- Guadagno, R. E., & Sagarin, B. J. (2010). Sex differences in jealousy: An evolutionary perspective on online infidelity. *Journal of Applied Social Psychology, 40*(10), 2636–2655.
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal 29*, 75–91.
- Guyz, D., Dickson-Swift, V., Kenny, A., & Threlkeld, G. (2015). Gadamerian philosophical hermeneutics as a useful methodological framework for the Delphi technique. *International Journal of Qualitative Studies on Health and Well-Being, 10*, 1–13.
- Hall-Kenyon, K. M., Bingham, G. E., & Korth, B. B. (2009). How do linguistically diverse students fare in full- and half-day kindergarten? Examining academic achievement, instructional quality, and attendance. *Early Education and Development, 20*, 25–52.
- Han, G. S., & Davies, C. (2006). Ethnicity, health and medical care: Towards a critical realist analysis of general practice in the Korean community in Sydney. *Ethnicity and Health, 11*(4), 409–430.
- Hannan, M., Happ, M. B., & Charron-Prochownik, D. (2009). Mothers' perspectives about reproductive health discussion with adolescent daughters with diabetes. *The Diabetes Educator, 35*(2), 265–273.
- Harris, G. T., Rice, M. E., & Cormier, C. A. (2013). Research and clinical scoring of the psychopathy checklist can show good agreement. *Criminal Justice and Behavior, 40*(11), 1349–1362.
- Haydon, T., Mancil, G. R., & Van Loan, C. (2009). Using opportunities to respond in a general education classroom: A case study. *Education and Treatment of Children, 32*(2), 267–278.
- Hedayat, A. S., Stufken, J., & Yang, M. (2006). Optimal and efficient crossover designs when subject effects are random. *Journal of the American Statistical Association, 101*, 1031–1038.
- Helmer, O. (1967). Analysis of the future: The Delphi method (No. RAND-P-3558). Santa Monica, CA: RAND Corporation.
- Hine, C. (2015). *Ethnography for the Internet: Embedded, embodied and every day*. New York, NY: Bloomsbury Academic.

- Hinkelman, K. (2012). *Design and analysis of experiments (Vol. 3): Special designs and applications*. Hoboken, NJ: Wiley.
- Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in early school years. *Journal of Educational Psychology, 101*(3), 662–670.
- Hirakata, P. (2009). Narratives of dissociation: Insights into the treatment of dissociation in individuals who were sexually abused as children. *Journal of Trauma & Dissociation, 10*(1), 297–314.
- Hlavinkova, L., Mentel, A., Kollarova, J., & Kristuffkova, Z. (2014). Effectiveness of a prevention campaign on HIV/AIDS knowledge among adolescents in Eastern Slovakia. *International Journal of Public Health, 59*(6), 905–911.
- Hofer, M., Burkhard, L., & Allemand, M. (2015). Age differences in emotion regulation during a distressing film scene. *Journal of Media Psychology: Theories, Methods, and Applications, 27*(2), 47–52.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*, 615–635.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge, UK: Cambridge University Press.
- Jaccard, J., & Jacoby, J. (2010). *Theory construction and model-building skills: A practical guide for social scientists*. New York, NY: Guilford Press.
- Jones, M. A., Stratten, G., Reilly, T., & Unnithan, V. B. (2004). A school-based survey of recurrent non-specific low-back pain prevalence and consequences in children. *Health Education Research, 19*(3), 284–289.
- Jones, S. R., & Hill, K. E. (2003). Understanding patterns of commitment: Student motivation for community service involvement. *Journal of Higher Education, 74*(5), 516–539.
- Kaldjian, L. C., Jones, E. W., Rosenthal, G. E., Tripp-Reimer, T., & Hillis, S. L. (2006). An empirically derived taxonomy of factors affecting physicians' willingness to disclose medical errors. *Journal of General Internal Medicine, 21*, 942–948.
- Kameoka, S., Yagi, J., Arai, Y., Nosaka, S., Saito, A., Miyake, W., . . . Asukai, N. (2015). Feasibility of trauma-focused cognitive behavioral therapy for traumatized children in Japan: A pilot study. *International Journal of Mental Health Systems, 9*, 26.
- Kartalova-O'Doherty, Y., & Doherty, D. T. (2008). Coping strategies and styles of family carers of persons with enduring mental illness: A mixed methods analysis. *Scandinavian Journal of Caring Sciences, 22*(1), 19–28.
- Kazdin, A. E. (2002). *Research design in clinical psychology* (4th ed.). Upper Saddle River, NJ: Pearson.
- Kazdin, A. E., Esveldt-Dawson, K., French, N. H., & Unis, A. S. (1987). Problem-solving skills training and relationship therapy in the treatment of antisocial child behavior. *Journal of Counseling and Clinical Psychology, 55*(1), 76–85.
- Keeney, S., McKenna, H., & Hasson, F. (2011). *The Delphi technique in nursing and health research*. Hoboken, NJ: Wiley.
- Keith, T. Z. (2014). *Multiple regression and beyond* (2nd ed.). Boston, MA: Pearson.
- Kemp, A., Quintana, D. S., Quinn, C., Hopkinson, P., & Harris, A. (2014). Major depressive disorder with melancholia displays robust alterations in resting state

- heart rate and variability: Implications for future morbidity and mortality. *Frontiers in Psychology*, 5. Retrieved from <http://dx.doi.org/10.3389/fpsyg.2014.01387>
- Kennedy, T. D., Burnett, K. F., & Edmonds, W. A. (2011). Intellectual, behavioral, and personality correlates of violent vs. non-violent juvenile offenders. *Aggressive Behavior*, 37(4), 315–325.
- Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Knapp, T. R., & Schafer, W. D. (2009). From gain score t to ANOVA F (and vice versa). *Practical Assessment, Research & Evaluation*, 14(6), 1–7.
- Kraemer, H. C., & Blasey, C. (2015). *How many subjects? Statistical power analysis in research* (2nd ed.). Thousand Oaks, CA: Sage.
- Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal*, 40(1), 281–310.
- Lachin, J. M., Matts, J. P., & Wei, L. J. (1988). Randomization in clinical trials: Conclusions and recommendations. *Controlled Clinical Trials*, 9, 365–374.
- Lapadat, J. C. (2004). Autobiographical memories of early language and literacy development. *Narrative Inquiry*, 14(1), 113–140.
- Lavrakas, P. J. (Ed.). (2009). *Encyclopedia of survey research methods*. Thousand Oaks, CA: Sage.
- Leake, M., & Lesik, S. A. (2007). Do remedial English programs impact first-year success in college? An illustration of the regression-discontinuity design. *International Journal of Research & Method in Education*, 30(1), 89–99.
- Lee, J. J., & Rubin, D. B. (2015). Valid randomization-based p -values for partially post hoc subgroup analyses. *Statistics in Medicine*. doi: 10.1002/sim.6531
- Lee, K. S., Osborne, R. E., Hayes, K. A., & Simoes, R. A. (2008). The effects of pacing on the academic testing performance of college students with ADHD: A mixed methods study. *Journal of Educational Computing Research*, 39(2), 123–141.
- Lee, S., Paek, J., Han, S., & Ryu, H. (in press). The utility of a Korean version of the REM sleep behavior disorder screening questionnaire in patients with obstructive sleep apnea. *Journal of the Neurological Sciences*.
- Lee, Y., Park, S., Kim, M., Son, C., & Lee, M. (2005). The effects of visual illustrations on learners' achievement and interest in PDA- (personal digital assistant) based learning. *Journal of Educational Computing Research*, 33(2), 173–187.
- Leech, N. L., Barrett, K. C., & Morgan, G. A. (2014). *SPSS for introductory and intermediate statistics: Use and interpretation* (7th ed.). London, England: Routledge.
- Lesik, S. A. (2006). Applying the regression-discontinuity design to infer causality with non-random assignment. *The Review of Higher Education*, 30(1), 1–19.
- Lesik, S. A. (2008). Studying the effectiveness of programs and initiatives in higher education using the regression-discontinuity design. *Handbook of Theory and Research*, 23, 277–297.
- Levy, P. S., & Lemeshow, S. (2009). *Sampling of populations: Methods and applications* (4th ed.). New York, NY: Wiley.

- Lieblich, A., Tuval-Mashiach, R., & Zilber, T. (1998). *Narrative research: Reading, analysis, and interpretation*. Thousand Oaks, CA: Sage.
- Lincoln, Y. S., & Guba, E. G. (2013). *The constructivist credo*. Walnut Creek, CA: Left Coast Press.
- Lincoln, Y. S., Linham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging influences., revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (4th ed.), pp. 191–215. Thousand Oaks, CA: Sage.
- Linden, A., Trochim, W. M. K., & Adams, J. L. (2006). Evaluating program effectiveness using the regression point displacement design. *Evaluation & the Health Professions, 29*(4), 407–423.
- Loh, J. (2013). Inquiry into issues of trustworthiness and quality in narrative studies: A perspective. *The Qualitative Report, 18*(65), 1–15.
- Lopez, E. N., Drobos, D. J., Thompson, K. J., & Brandon, T. H. (2008). Effects of body image challenge on smoking motivation among college females. *Health Psychology, 27*(3), 243–251.
- Luytena, H., Tymms, P., & Jones, P. (2009). Assessing school effects without controlling for prior achievement? *School Effectiveness and School Improvement, 20*(2), 145–165.
- Lyon, A. R., Gershenson, R. A., Farahmand, F. K., Thaxter, P. J., Behling, S., & Budd, K. S. (2009). Effectiveness of teacher-child interaction training (TCIT) in a pre-school setting. *Behavior Modification, 33*(6), 855–884.
- Madison, D. S. (2011). *Critical ethnography: Methods, ethics, and performance* (2nd ed.). Thousand Oaks, CA: Sage.
- Maguire, P. (1987). *Doing participatory research: A feminist approach*. Amherst, MA: Center for International Education.
- Makagon, D., & Neumann, M. (2008). *Recording culture*. Thousand Oaks, CA: Sage.
- Malcolm, T. (2010). The curious case of case study: A viewpoint. *International Journal of Social Research Methodology: Theory and Practice, 13*(4), 329–339.
- Mancil, G. R., Haydon, T., & Whitby, P. (2009). Differential effects of paper and computer-assisted Social Stories™ on inappropriate behavior in children with autism. *Focus on Autism and Other Developmental Disabilities, 24*(4), 205–215.
- Manley, R. A., & Zinser, R. (2012). A Delphi study to update CTE teacher competencies. *Education & Training, 54*(6), 488–503.
- Manolov, R., Solanas, A., Bulté, I., & Onghena, P. (2010). Data-division-specific robustness and power of randomization tests for ABAB designs. *Journal of Experimental Education, 78*, 191–214.
- May, H., & Supovitz, J. A. (2006). Capturing the cumulative effects of school reform: An 11-year study of the impacts of America's choice on student achievement. *Educational Evaluation and Policy Analysis, 28*(3), 231–257.
- McAdams, D. P., Josselson, R., & Lieblich, A. (2006). *Identity and story: Creating self in narrative*. Washington, DC: American Psychological Association.
- McCarthy, A. M., & Tucker, M. L. (2002). Encouraging community service through service learning. *Journal of Management Education, 26*(6), 629–647.

- McNemar, Q. (2007). On the use of Latin squares in psychology. *Psychological Bulletin*, 48(5), 398–401.
- Meijer, J., & Oostdam, R. (2007). Test anxiety and intelligence testing: A closer examination of the stage-fright hypothesis and the influence of stressful instruction. *Anxiety, Stress, and Coping*, 20(1), 77–91.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2012). *Applied multivariate research: Design and interpretation* (2nd ed.). Thousand Oaks, CA: Sage.
- Moerer-Urdahl, T., & Creswell, J. W. (2004). Using transcendental phenomenology to explore the “ripple effect” in a leadership mentoring program. *International Journal of Qualitative Methods*, 3(2), 19–35.
- Morgan, B. J. (2001). Evaluation of an educational intervention for military tobacco users. *Military Medicine*, 166(12), 1094–1098.
- Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, 40, 120–123.
- Morse, J. M., & Niehaus, L. (2009). *Mixed methods design: Principles and procedures*. Walnut Creek, CA: Left Coast Press.
- Mundia, L. (2011). Social desirability, non-response bias and reliability in long self-report measure: Illustrations from the MMPI-2 administered to Brunei student teachers. *Educational Psychology*, 31(2), 207–224.
- Nagel, T., Robinson, G., Condon, J., & Trauer, T. (2009). Approach to treatment of mental illness and substance dependence in remote indigenous communities: Results of a mixed methods study. *Australian Journal of Rural Health*, 17(4), 174–182.
- Nastasi, B. K., Hitchcock, J., Sarkar, S., Burkholder, G., Varjas, K., & Jayasena, A. (2007). Mixed methods in intervention research: Theory to adaptation. *Journal of Mixed Methods Research*, 1(2), 164–183.
- O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin*, 97(2), 316–333.
- Oltmanns, T. F., Martin, M. T., Neale, J. M., & Davison, G. C. (2014). *Case studies in abnormal psychology* (10th ed.). San Francisco, CA: Wiley.
- Pace, L. A. (2011). *Statistical analysis using Excel 2007*. Upper Saddle River, NJ: Prentice Hall.
- Palinkas, L. A., Horwitz, S. A., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2014). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administrative Policy and Mental Health*. doi:10.1007/s10488-013-0528-y
- Park, Y. S., Riddle, J., & Tekian, A. (2014). Validity evidence of resident competency ratings and the identification of problem residents. *Medical Education*, 48(6), 614–622.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Perilloux, C., Cloud, J. M., & Buss, D. M. (2013). Women's physical attractiveness and short-term mating strategies. *Personality and Individual Differences*, 54, 490–495.

- Plano Clark, V. L. (2005). Cross-disciplinary analysis of the use of mixed methods in physics education research, counseling psychology, and primary care. *Dissertation Abstracts International*, 66, 02A.
- Polkinghorne, D. E. (1988). *Narrative knowing and the human sciences*. Albany: State University of New York Press.
- Pollio, H. R., Henley, T., & Thompson, C. B. (1997). *The phenomenology of everyday life*. Cambridge, UK: Cambridge University Press.
- Prinzie, P., & Onghena, P. (2005). Cohort sequential design. In B. S. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 319–322). West Sussex, UK: Wiley-Blackwell.
- Probst, T. M. (2003). Exploring employee outcomes of organizational restructuring. *Group & Organization Management*, 28(3), 416–439.
- Purser, G. (2009). The dignity of job-seeking men: Boundary work among immigrant day laborers. *Journal of Contemporary Ethnography*, 38(1), 117–139.
- Ragin, C. C. (1997). Turning the tables: How case-oriented research challenges variable-oriented research. *Comparative Social Research*, 16, 27–42.
- Randall, R. R., & Barcikowski, R. S. (1987, April). *Sphericity tests and repeated measures data*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.
- Reese, H. W. (1997). Counterbalancing and other uses of repeated-measures Latin-square designs: Analysis and interpretations. *Journal of Experimental Child Psychology*, 64, 137–158.
- Reynolds, L. M., McCambridge, S. A., Bissett, I. P., & Consedine, N. S. (2014). Trait and state disgust: An experimental investigation of disgust and avoidance in colorectal cancer decision scenarios. *Health Psychology*, 33(12), 1495–1506.
- Richardson, J. B. (2009). Men do matter: Ethnographic insights on the socially supportive role of the African American uncle in the lives of inner-city African American male youth. *Journal of Family Issues*, 30(8), 1041–1069.
- Riessman, C. K. (2007). *Narrative methods for the human sciences*. Thousand Oaks, CA: Sage.
- Rihoux, B., & Lobe, B. (2009). The case for qualitative comparative analysis (QCA): Adding leverage for thick cross-case comparison. In D. Byrne & C. C. Ragin (Eds.), *The Sage handbook of case-based methods* (pp. 222–242). Thousand Oaks, CA: Sage.
- Rimondini, M., Del Piccolo, L., Goss, C., Mazzi, M., Paccaloni, M., & Zimmerman, C. (2010). The evaluation of training in patient-centered interviewing skills for psychiatric residents. *Psychological Medicine*, 40, 467–476.
- Rubin, D. B. (2006). *Matched sampling for causal effects*. Cambridge, UK: Cambridge University Press.
- Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*, 26(1), 20–36.
- Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2(3), 808–840.

- Ryan, T. P. (2007). *Modern experimental design*. Hoboken, NJ: Wiley.
- Ryndak, D. L., Storch, J. F., & Hoppey, D. (2008). One family's perspective of their experiences with school and district personnel over time related to inclusive educational services for a family member with significant disabilities. *International Journal of Whole Schooling*, 4(2), 29–51.
- Sawilowsky, S., Kelley, D. L., Blair, R. C., & Markman, B. S. (1994). Meta-analysis and the Solomon four-group design. *Journal of Experimental Education*, 62, 361–376.
- Schoenfeld, N. A., & Mathur, S. R. (2009). Effects of cognitive-behavioral intervention on the school performance of students with emotional or behavioral disorders and anxiety. *Behavioral Disorders*, 34(4), 184–195.
- Schultz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Research Methods & Reporting*, 340, 698–702.
- Seifert, T. A., Goodman, K., King, P. M., & Baxter Magolda, M. B. (2010). Using mixed methods to study first-year college impact on liberal arts learning outcomes. *Journal of Mixed Methods Research*, 4(3), 248–267.
- Shadish, W. R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods*, 15, 3–17.
- Shadish, W. R. (2014a). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology*, 52, 109–122.
- Shadish, W. R. (2014b). Statistical analyses of single-case designs: The shape of things to come. *Current Directions in Psychological Science*, 23, 139–146. Retrieved from <http://faculty.ucmerced.edu/wshadish/software/software-meta-analysis-single-case-design>
- Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random to nonrandom assignment. *Journal of the American Statistical Association*, 103, 1334–1344.
- Shadish, W. R., & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, 60, 607–629.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shin, K. R. (2002). Using hermeneutic phenomenology to elicit a sense of body at mid-life. *International Journal of Qualitative Methods*, 1(2), 39–50.
- Sidani, Y. M. (2007). Perceptions of leader transformational ability: The role of leader speech and follower self-esteem. *Journal of Management Development*, 26(8), 710–722.
- Singer, J. A. (1997). *Message in a bottle: Stories of men and addiction*. New York, NY: Free Press.
- Singer, J. A., & Bonalume, L. (2010a). Toward the scientific study of autobiographical memory narratives in psychotherapy. *Pragmatic Case Studies in Psychotherapy*, 6, 215–222.

- Singer, J. A., & Bonalume, L. (2010b). Autobiographical memory narratives in psychotherapy: A coding system applied to the case of Cynthia. *Pragmatic Case Studies in Psychotherapy*, 6, 134–188.
- Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretive phenomenological analysis: Theory, method, and research*. Thousand Oaks, CA: Sage.
- Smith, M. E. (2007). Self-deception among men who are mandated to attend a batterer intervention program. *Perspectives in Psychiatric Care*, 43(4), 193–203.
- Solomon, R. L. (1949). An extension of the control group design. *Psychological Bulletin*, 46, 137–150.
- Spector, P. E. (1981). *Research designs*. Beverly Hills, CA: Sage.
- Spector, P. E. (1997). *Job satisfaction: Application, assessment, causes, and consequences*. Thousand Oaks, CA: Sage.
- Spradley, J. P. (1979). *The ethnographic interview*. New York, NY: Holt, Rinehart & Winston.
- Spradley, J. P. (1980). *Participant observation*. New York, NY: Holt, Rinehart & Winston.
- Steinfeld, C., Ellison, N. B., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, 29, 434–445.
- Stern, S. E., Mullennix, J. W., & Wilson, S. J. (2002). Effects of perceived disability on persuasiveness of computer-synthesized speech. *Journal of Applied Psychology*, 87(2), 411–417.
- Stewart, A. L., King, A. C., & Haskell, W. L. (1993). Endurance exercise and health-related quality of life in 50–65 year-old adults. *The Gerontologist*, 33(6), 782–789.
- Steyn, R. (2009). Re-designing the Solomon four-group: Can we improve on this exemplary model? *Design Principles and Practices: An International Journal*, 3(1), 1833–1874.
- Stringer, E. T. (2013). *Action research* (4th ed.). Thousand Oaks, CA: Sage.
- Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, & applications* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Tashakkori, A., & Teddlie, C. B. (2002). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. B. (2010a). *SAGE handbook of mixed methods in social & behavioral research* (2nd ed.). Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. B. (2010b). Putting the human back into “human research methodology”: The researcher in mixed methods research. *Journal of Mixed Methods Research*, 4(4), 271–277.
- Tate, R., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., Shadish, W. R., . . . Vohra, S. (in preparation). The single-case reporting guideline in behavioural interventions (SCRIBE): Explanation and elaboration. *Archives of Scientific Psychology*.
- Thomas, J. (1993). *Doing critical ethnography*. Newbury Park, CA: Sage.
- Thompson, K. J., & Oliver, B. E. (2012). Response rates in business surveys: Going beyond the usual performance measure. *Journal of Official Statistics*, 28(2), 221–237.

- Trochim, W. (2001). Regression-discontinuity design. In N. J. Smelser, J. D. Wright, & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (Vol. 19, pp. 12940–12945). North-Holland, Amsterdam: Pergamon.
- Trochim, W., & Campbell, D. T. (1996). *The regression point displacement design for evaluating community-based pilot programs and demonstration projects*. Unpublished manuscript. Retrieved from <http://www.socialresearchmethods.net/research/RPD/RPD.pdf>
- Trochim, W., & Cappelleri, J. C. (1992). Cutoff assignment strategies for enhancing randomized clinical trials. *Controlled Clinical Trials*, *13*, 190–212.
- Vagle, M. D. (2014). *Crafting phenomenological research*. Walnut Creek, CA: Left Coast Press.
- van Maanen, J. (1988). *Tales of the field: On writing ethnography*. Chicago, IL: University of Chicago Press.
- van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. Albany: The State University of New York.
- van Manen, M. (2014). *Phenomenology of practice: Meaning-giving methods in phenomenological research and writing*. Walnut Creek, CA: Left Coast Press.
- Varcoe, C., Browne, A. J., Wong, S., & Smye, V. L. (2009). Harms and benefits: Collecting ethnicity data in a clinical context. *Social Science & Medicine*, *68*, 1659–1666.
- Villalta-Gil, V., Roca, M., Gonzales, N., Domènec, E., Cuca, Escanilla, A., . . . Haro, J. M. (2009). Dog-assisted therapy in the treatment of chronic schizophrenia inpatients. *Antrozoös*, *22*(2), 149–159.
- Viswanathan, M. (2005). *Measurement error and research design*. Thousand Oaks, CA: Sage.
- Vogt, W. P. (2005). *Dictionary of statistics and methodology: A non-technical guide for the social sciences* (3rd ed.). Thousand Oaks, CA: Sage.
- von Eckartsberg, R. (1997). Introducing existential-phenomenological psychology. In R. Valle (Ed.), *Phenomenological inquiry in psychology: Existential and transpersonal dimensions* (pp. 3–20). New York, NY: Plenum.
- Walker, C. O., & Greene, B. A. (2009). The relations between student motivational beliefs and cognitive engagement in high school. *Journal of Educational Research*, *102*(6), 463–471.
- Walton Braver, M. C., & Braver, S. L. (1988). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Psychological Bulletin*, *104*(1), 150–154.
- Wertz, F. J., Charmaz, K., McMullen, L. M., Josselson, R., Anderson, R., & Emalinda, M. (2011). *Five ways of doing qualitative analysis: Phenomenological psychology, grounded theory, discourse analysis, narrative research, and intuitive inquiry*. New York, NY: Guilford Press.
- White, P. (2009). *Developing research questions: A guide for social scientists*. New York, NY: Palgrave Macmillan.
- Whitman, D. S., Van Rooy, D. L., Viswesvaran, C., & Alonso, A. (2008). The susceptibility of a mixed model measure of emotional intelligence to faking: A Solomon four-group design. *Psychology Science*, *50*(1), 44–63.

- Williams, E. N., & Morrow, S. L. (2009). Achieving trustworthiness in qualitative research: A pan-paradigmatic perspective. *Psychotherapy Research, 19*(4–5), 576–582.
- Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*. Thousand Oaks, CA: Sage.
- Wyatt, T. H., & Hauenstein, E. J. (2008). Pilot test Okay With Asthma™: An online asthma intervention for school-age children. *Journal of School Nursing, 24*(3), 145–150.
- Yin, R. K. (Ed.). (2004). *The case study anthology*. Thousand Oaks, CA: Sage.
- Yin, R. K. (2012). *Applications of case study research* (3rd ed.). Thousand Oaks, CA: Sage.
- Yin, R. K. (2013). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.
- Young, P., & Miller-Smith, K. (2006). Effects of state mandated policy (site-based councils) and of potential role incumbents on teacher screening decisions in high and low performing schools. *Education Policy and Analysis Archives, 14*(7), 1–21.
- Zimmerman, D. W., & Zumbo, B. D. (2009). Hazards in choosing between pooled and separate-variances *t* tests. *Psicológica, 30*, 371–390.
- Zoffmann, V., & Kirkevold, M. (2007). Relationships and their potential for change developed in difficult type 1 diabetes. *Qualitative Health Research, 17*(5), 625–638.
- Zolotor, A. J., Runyan, D. K., Dunne, M. P., Jain, D., Petrus, H. R., Ramirez, C., . . . Isaeva, O. (2009). ISPCAN Child Abuse Screening Tool Children's Version (ICAST-C): Instrument development and multi-national pilot testing. *Child Abuse & Neglect, 33*(11), 833–841.
- Zydney, J. M. (2008). Cognitive tools for scaffolding students defining an ill-structured problem. *Journal of Educational Computing Research, 38*(4), 353–385.

Author Index

- Abrami, P., 40
Abrams, L. S., 186
Acee, T. W., 65
Adams, J. L., 219, 225
Allemand, M., 236
Alonso, A., 97
Altman, D. G., 32, 337
Anderson, R., 144
Anderson, R. B., 126
Aydin, E., 130
- Baker, M., 32
Baltes, P. B., 53
Barcikowski, R. S., 254
Barrett, K. C., 220
Basadur, M., 67
Baxter Magolda, M. B., 181
Baylor, A. L., 85
Bazeley, P., 220
Becke, S., 194
Behling, S., 113
Berman, H., 166
Bernard, R. S., 107
Bingham, G. E., 183
Birks, M., 145
Bissett, I. P., 236
Blair, R. C., 93
Blasey, C., 220, 241
Bohannon, J., 32
Bonalume, L., 143, 161
Boonyasai, R. T., 239
Borg, W. R., 144
Bortree, D. S., 240
Box, G. E. P., 63
Brandon, T. H., 81
Braver, S. L., 93
Browne, A. J., 156
Bryan, T., 37
Bryant, B. R., 53
Bryant, D. P., 53
- Buchner, A., 220
Buckles, D. J., 212, 214
Budd, K. S., 113
Bulté, I., 104
Burgess, G., 69
Burkhard, L., 236
Burnett, K. F., 239
Burstein, K., 37
Burwitz, L., 69
Buss, D. M., 242–243
Butler-Kisber, L., 221
- Campbell, D. T., 4, 12, 15, 30, 31, 37, 52, 219, 221, 225
Cappelleri, J. C., 53
Cernin, P. A., 66
Chakraborti, C., 239
Chao, P., 37
Chapin, M. H., 122
Chapman, Y., 145
Charmaz, K., 144, 147, 150f
Charron-Prochownik, D., 199
Chavez, M. M., 53
Chenail, R., 4, 142, 328–330
Cheng, S., 193
Chevalier, J. M., 212, 214
Cifuentes, L., 46
Clandinin, D. J., 162
Clark, M. H., 34
Cloud, J. M., 242–243
Cohen, J., 237
Cohen, L. L., 107
Collins, K. M. T., 179
Comaskey, E. M., 40
Comier, C. A., 243
Condon, J., 205
Connelly, F. M., 162
Consedine, N. S., 236
Cook, T. D., 4, 12, 17, 30, 31, 37, 52, 221
Coppola, L., 244
Corbin, J., 145, 147

- Coryn, C. L. S., 220
 Creswell, J. W., 22, 143–144, 153, 157*f*, 172, 174*n*, 177,
 178, 179, 180, 201, 210*n*, 218–219, 229
- Dasgupta, T., 77
 Davies, C., 147
 Del Piccolo, L., 70
 Dennis, J. K., 46
 Denzin, N. K., 144
 DeVellis, R. F., 4
 Dickson-Swift, V., 221
 Dixon, M. R., 105
 Doherty, D. T., 204
 Doherty, M. E., 126
 Domenec, E., 47
 Drechsler, J., 135
 Drobes, D. J., 81
 Duan, N., 180
 Dunne, M. P., 203
- Edmonds, W. A., xxvi, 239
 Edmunds, K. A., 166
 Elliott, M. S., 187–188
 Ellison, N. B., 137
 Emalinda, M., 144
 Erdfelder, E., 220
 Erdogan, Y., 130
 Ergul, C., 37
 Escanilla, A., 47
 Esveldt-Dawson, K., 39
- Farahmand, F. K., 113
 Faul, F., 220
 Fetterman, D. M., 154
 Finchman, J. E., 11*f*
 Fink, A. G., 135
 Fletcher, A., 213
 Flores, M. M., 110
 Flowers, P., 170
 Forchuk, C., 166, 193
 Francis, K., 145
 French, N. H., 39
 Friedrich, J. C., 126
- Gall, J. P., 144
 Gall, M. D., 144
 Gamst, G., 129
 Ganz, J. B., 110
 Gast, D. L., 106
 Geertz, C., 153
 Gershenson, R. A., 113
 Gersten, R., 53
- Ghazarian, S. R., 131
 Gibney, C., 193
 Gier, V. S., 83
 Giuffre, M., 121
 Glaser, B. G., 145, 147, 148*f*
 Glass, V. G., 61
 Glombiewski, J. A., 238
 Golafshani, N., 4, 142
 Goodman, K., 181
 Goss, C., 70
 Gottman, J. M., 61
 Graen, G. B., 67
 Green, C. A., 180
 Green, M., 46
 Green, S., 220
 Greene, B. A., 129
 Grogam, S., 69
 Guadagno, R. E., 89
 Guarino, A. J., 129
 Guba, E. G., 4, 144, 324
 Guyz, D., 221
- Haldenby, A., 166
 Hall-Kenyon, K. M., 183
 Han, G. S., 147
 Hannan, M., 199
 Happ, M. B., 199
 Haro, J. M., 47
 Harris, A., 237
 Harris, G. T., 243
 Haskell, W. L., 48
 Hasson, F., 213
 Hausenstein, E. J., 64
 Haydon, T., 106, 109
 Hayes, K. A., 198
 Hedayat, A. S., 61
 Heidegger, M., 169
 Helmer, O., 213
 Henley, T., 170–171
 Hill, K. E., 150
 Hillis, S. L., 184
 Hine, C., 152
 Hinkelman, K., 63
 Hinnant, J. B., 131
 Hirakata, P., 164
 Hlavinkova, L., 238
 Hoagwood, K., 180
 Hofer, M., 236
 Holbert, D., 122
 Holsti, O., 240
 Hopkinson, P., 237
 Hoppey, D., 174–175

- Horwitz, S. A., 180
Husserl, E., 168
- Imbens, G. W., 16, 53, 54, 118, 129
Isaeva, O., 203
- Jaccard, J., 3
Jackson, K., 220
Jacobson, J., 3
Jain, D., 203
Jones, E. W., 184
Jones, M. A., 136
Jones, P., 51
Jones, S. R., 150
Josselson, R., 144, 161
- Kabaca, T., 130
Kaiser, M. K., 254
Kaldjian, L. C., 184
Kameoka, S., 238
Kartalova-O'Doherty, Y., 204
Kazdin, A. E., 39, 75
Keeney, S., 213
Keith, T. Z., 129
Kelley, D. L., 93
Kemp, A., 237
Kennedy, T. D., xxvi-xxvii, 239
Kenny, A., 221
Kern, D. E., 239
Kim, M., 41
Kim, S., 85
King, A. C., 48
King, P. M., 181
Kirkevold, M., 149
Kline, R. B., 126
Knapp, T. R., 253
Kollarova, J., 238
Korth, B. B., 183
Kraemer, H. C., 220, 241
Kramarski, B., 42
Kreiner, D. S., 83
Kristufkova, Z., 238
- Lachin, J. M., 17*n*
Lampe, C., 137
Lang, A. G., 220
Lapadat, J. C., 162
Larkin, M., 170
Lavrakas, P. J., 135
Leake, M., 55
Ledford, J. R., 106
Lee, J. J., 31
Lee, K. S., 198
- Lee, M., 41
Lee, S., 240
Lee, Y., 41
Leech, N. L., 220
Lemeshow, S., 21
Lemieux, T., 53, 54
Lesik, S. A., 55
Levy, P. S., 21
Lichtenberg, P. A., 66
Lieblich, A., 161, 162, 165
Lincoln, Y. S., 4, 144
Linden, A., 219, 225
Linham, S. A., 144
Lobe, B., 142
Loh, J., 4
Lopez, E. N., 81
Lopez, R., 166
Luytena, H., 51
Lyon, A. R., 113
- Madison, D. S., 153, 154
Makagon, D., 154
Malcolm, T., 143
Malla, A., 193
Mancil, G. R., 106, 109
Manley, R. A., 213
Manolov, R., 104
Marchildon, G., 213
Markman, B. S., 93
Martin, M., 193
Mathur, S. R., 111
Matts, J. P., 17*n*
Maxin, J., 194
May, H., 71
Mazzi, M., 70
McAdams, D., 161
McCambridge, S. A., 236
McCarthy, A. M., 98
McKenna, H., 213
McLean, T., 193
McMullen, L. M., 144
McNemar, Q., 63
Meijer, J., 72
Mentel, A., 238
Mevarach, Z. R., 42
Meyers, L. S., 129
Miller-Smith, K., 87
Moerer-Urdahl, T., 172
Moffett, K., 107
Moher, D., 32, 337
Morgan, B. J., 122
Morgan, G. A., 220
Morrow, S. L., 4

- Morse, J. M., 179*n*, 180
 Mulcachy, G. A., 166
 Mullennix, J. W., 77
 Mundia, L., 135
- Nagel, T., 205
 Nastasi, B. K., 179*n*
 Natz-Gonzalez, A., 83
 Neumann, M., 154
 Niehaus, L., 179*n*, 180
 Norman, R., 193
- O'Brien, M., 131
 O'Brien, R. G., 254
 O'Cathain, A., 179
 Oliver, B. E., 135
 Oltmanns, T. F., 189
 Onghena, P., 104, 135
 Oostdam, R., 72
 Osborne, R. E., 198
- Paccaloni, M., 70
 Pace, L. A., 220
 Paek, J., 240
 Palinkas, L. A., 20, 180
 Park, S., 41
 Park, Y. S., 239
 Patton, M. Q., 20
 Perilloux, C., 242–243
 Petrus, H. R., 203
 Pillai, N. S., 77
 Plano-Clark, V. L., 177, 178, 179, 180, 201, 219
 Polkinghorne, D. F., 106
 Pollio, H. R., 170–171
 Prinzie, P., 135
 Probst, T. M., 95
 Purser, G., 154
- Quinn, C., 237
 Quintana, D. S., 237
- Ragin, C. C., 142
 Ramirez, C., 203
 Randall, R. R., 254
 Reese, H. W., 63
 Reilly, T., 136
 Reynolds, L. M., 236
 Rice, M. E., 243
 Richardson, J. B., 157
 Riddle, J., 239
 Riessman, C. K., 160, 162
 Rihoux, B., 142
 Rimondini, M., 70
- Robinson, G., 205
 Roca, M., 47
 Rosenthal, G. E., 184
 Rubin, D. B., 16, 17, 31, 77, 118, 129
 Runyan, D. K., 203
 Ryan, T. P., 68, 77
 Ryndak, D. L., 174–175
 Ryu, H., 240
- Sagarin, B. J., 89
 Salkind, N. J., 220
 Sangalang, C., 186
 Savage, R. S., 40
 Sawilowsky, S., 93
 Scammacca, N., 53
 Scandura, T. A., 67
 Schafer, W. D., 253
 Schoenfeld, N. A., 111
 Schultz, K. F., 32, 337
 Seifert, T. A., 181
 Seltzer, T., 240
 Shadish, W. R., 4, 12, 13, 16, 31, 34, 36, 37, 52, 68, 105, 106, 172
 Shannon, S. K. S., 186
 Shin, K. R., 173–174
 Sidani, Y. M., 79–80
 Simoes, R. A., 198
 Singer, J., 143, 161
 Smelser, N. J., 53
 Smith, J. A., 170
 Smith, M. E., 170
 Smye, V. L., 156
 Solanas, A., 104
 Solomon, R. L., 93
 Son, C., 41
 Spector, P. E., 64, 121
 Spradley, J. P., 153
 Stanley, J. C., 15
 Steiner, P. M., 17, 34
 Steinfield, C., 137
 Stern, S. F., 77
 Stewart, A. L., 48
 Steyn, R., 94
 Storch, J. F., 174–175
 Stratten, G., 136
 Strauss, A., 145, 147, 148*f*
 Stringer, E. T., 212, 214
 Stufflebeam, D. L., 220
 Stufken, J., 61
 Supovitz, J. A., 71
- Tashakkori, A., 177, 179, 180, 181, 208, 231
 Tate, R., 103

- Teddlie, C. B., 177, 179, 180,
 181, 208, 231
 Tekian, A., 239
 Thaxter, P. J., 113
 Thomas, J., 153
 Thompson, C. B., 170–171
 Thompson, K. J., 81, 135
 Threlkeld, G., 221
 Trauer, T., 205
 Tripp-Reimer, T., 184
 Trochim, W., 53, 219, 225
 Tucker, M. L., 98
 Tuval-Mashiach, R., 162, 165
 Tymms, P., 51

 Unis, A. S., 39
 Unnithan, V. B., 136

 Vagle, M. D., 170, 172
 Van Loan, C., 106
 Van Maanen, J., 152–153, 154,
 161–162
 van Manen, M., 170, 172, 174
 Van Rooy, D. L., 97
 Varcoe, C., 156
 Villalta-Gil, V., 47
 Viswanathan, M., 4
 Viswesvaran, C., 97
 Vogt, W. P., 75
 von Eckartsberg, R., 169

 Walker, C. O., 129
 Walton Braver, M. C., 93
 Wei, L. J., 17*n*
 Weinstein, C. E., 65
 Wertz, F. J., 143, 144
 Whitby, P., 109
 White, P., 28
 Whitman, D. S., 97
 Wilkinson, R., 194
 Williams, D. I., 187–188
 Williams, E. N., 4
 Wilson, S. J., 77
 Wilson, V. L., 61
 Wisdom, J. P., 180
 Wolcott, H. F., 175
 Wong, S., 156
 Wright, J. D., 53
 Wright, S. M., 239
 Wyatt, T. H., 64

 Yang, M., 61
 Yin, R. K., 143–144, 193*n*, 210*n*, 229
 Young, P., 87

 Zilber, T., 162, 165
 Zimmerman, C., 70
 Zinser, R., 213
 Zoffmann, V., 149
 Zolotor, A. J., 203
 Zydney, J. M., 191

Subject Index

- A-B-A-B-C-B-C single-case design, 108–109
- A-B-A-B single-case design, 107–108
- A-B-A concurrent design, 209
- A-B-A nested design, 210
- A-B-A sequential design, 210
- A-B-A single-case design, 104, 106–107
 - mixed methods, 209–210
- A-B single-case design, 104
- Accelerated longitudinal design, 135
- Action research approaches, 212–214
 - design notations, 215*t*
 - embedded designs, 214–216
 - participatory action research, 213
 - reference resources, 214
- Alpha value (α), 241–242
- Alternating-treatment design, 105
- Analysis of covariance (ANCOVA) design, 35, 127, 252–253, 256–261, 306–314
- Analysis of variance (ANOVA), 127, 253
 - between-subjects example, 285–292
 - independent variable in, 5
 - multivariate (MANOVA), 254
 - repeated measures (RM ANOVA), 128, 254, 269–282
 - two-way, 128, 253
 - within-subjects example, 278–282
- APA style, 232, 244, 246
- Approach
 - appropriate designs for research questions, 23–26
 - qualitative research terminology, xviii
 - quantitative research terminology, xvii
- Area sampling, 20
- Arts-based approaches, 221
- Assessment reactivity, 10*t*
- Assumptions violations, threats to statistical conclusion validity, 11*t*
- Attention control group, 18
- Attention threat to construct validity, 10*t*
- Attrition, 8*t*
- Audio data analysis, 322
- Autoethnography, 221
- Axial coding, 148, 326–327
- Baselines, 105
 - multiple-baselines design, 105, 111–114
- Behavior research, 142
- Between-subjects approaches, 35
 - data analysis, 252–253
 - data analysis, examples, 255–278, 284–305, 315–320
 - factorial designs, 76–80, 82–89
 - k*-factor designs, 36–37
 - less common designs, 223–225
 - matched pairs assignment, 16
 - mixed-subjects, 59, 76, 89, 269
 - posttest designs, 43–49
 - pretest and posttest designs, 35–43
 - random control trial (RCT) designs, 32
 - regression-discontinuity approach, 51. *See also* Regression-discontinuity (RD) approach
 - Solomon *N*-group design, 93–101
 - See also* Pretest and posttest designs; *specific designs*
- Binary logistic regression, 128
- Blocking, 15
- Blocking factor interactions, 227*n*
- Case study designs, 143–144, 229–230
 - embedded designs, 190, 193*f*, 194–195, 229–230
 - ethnographic perspective, 153–154, 157–158, 229
 - mixed methods, 208–211
 - phenomenological perspective, 169–170, 174–175
 - See also* Single-case approaches
- Case-to-case generalization, 142
- Causal-comparative research, 120
- Cause and effect relationships
 - checklist, 333
 - conditions for establishing, 30–31, 221
 - nonexperimental research and, 118, 126
 - qualitative comparative analysis and, 142
- Central research question, 22–23, 26–27
- Changing criterion design, 105, 109–110
- Circularity assumption, 254
- Coding, 146, 148, 325–327
- Cohen's *d*, 237, 266, 318

- Cohen's kappa, 240
- Cohort-based data collection, 135
- Cohort controls, 18
- Cohort matching, 45
- Cohort-sequential design, 135
- Community-based action research, 213
- Comparison research questions, 22, 25–26
- Compensatory equalization, 11*t*
- Compensatory rivalry, 11*t*
- Conceptual definition, 6
- Concurrent nested mixed methods design, 191*n*
- Concurrent triangulation design, 181, 184*n*
- Confirmability, 324
- CONSORT 2010, 334–337
- Constant comparative method, 150
- Construct confounding, 10*t*
- Constructivist design, 146, 150–151, 218
- Construct validity, 9, 240–241
 - Solomon *N*-group design, 93
 - survey research, 133, 134
 - threats to, 10–11*t*, 134
- Content validity, 241
- Control, 13–14, 31
 - checklist, 333
 - quasi-experimental designs and, 33
 - statistical procedures and nonexperimental research, 118, 125
- Control groups, 18
 - less common designs, 222
 - wait-list, 18, 59
 - See also* Historical control group
- Convenience sampling, 20
- Convergence model, 181
- Convergent-parallel approach, 181
 - data-transformation design, 181, 184–185
 - data-validation design, 181, 185–186
 - multilevel design, 181, 187–188
 - parallel-databases design, 181, 183–184
- Correlational analysis, 125
 - explanatory design, 126
 - predictive design, 126–127
 - See also* Observational approach
- Correlational design, embedded, 190, 192–194
- Correlation statistics, 238–239, 243–244, 264, 298–299, 317, 319
- Counterbalancing, 17, 57, 58, 254
- Covariate, 36
- Credibility, 324
- Criterion-related validity, 241
- Criterion sampling, 20
- Criterion variables, 5–6, 127, 129
- Critical design, 153, 156–157, 161–162, 165*f*, 166
- Critical effect size, 241–242
- Cronbach's alpha, 239–240
- Crossover designs, 36, 59, 61, 62, 68
 - 2-factor, 68–70
 - 3-factor, 226, 252, 278–284
- Cross-sectional survey design, 135, 136–137
- Cues, threat to construct validity, 10*t*
- Culture, 152
- Cutoff score, 16, 51–52
- Data analytics, 220. *See also* Qualitative data analysis; Statistical analysis
- Data presentation for qualitative research, 330–331
- Data-transformation design, 181, 184–185
- Data trustworthiness, 4
- Data-validation design, 181, 185–186
- Deductive reasoning, 30, 218
- Degrees of freedom (df), 243
- Delphi technique, 213, 221
- Dependability, 324
- Dependent samples (paired samples) *t* test, 293, 301–302
- Dependent variables (DVs), 5–6
 - outcome variable for nonexperimental research, 5, 127, 129
- Descriptive design, 133, 161, 162–164
- Descriptive research questions, 22, 24–25
- Descriptive statistics, reporting preliminary findings, 242–244
- Design. *See* Research design
- Dialogic listening skills, 160
- Diffusion threat to internal validity, 8*t*
- Dispersion statistics, 242
- Double pretest and posttest control group design, 223, 225
- Double pretest design, 33*f*
- Dramatic data presentation, 331
- Effect size
 - converting *t*- to *r*-value, 302
 - critical, 241–242
 - ex post facto design (independent samples *t* test), 317–319
 - pretest and posttest designs (independent samples *t* test), 266
 - reporting, 237–239
 - statistical power and, 241
 - threats to statistical conclusion validity, 12*t*
- Eidetic reduction, 169
- Elimination and control, 14
- Embedded approach, 144, 189–190
 - action research and, 214–216
 - case study design, 190, 193*f*, 194–195, 229
 - embedded-correlational design, 190, 192*f*, 193–194
 - embedded-experiment design, 190, 191–193
- Emergent design in mixed methods, 178
- Emerging design in grounded theory, 146, 148, 149

- Epoché*, 173
- Error rate problem, 12*t*
- Ethnographic perspective, 152
 - autoethnography, 221
 - case study design, 153–154, 157–158, 229
 - critical design, 153, 156–157
 - embedded designs, 189
 - realist design, 152–153, 154–156
 - reference resources, 154
 - when to use, 154
- Evaluation, 213, 219–220
- Existential design, 168–169, 170–171
- Experimental research, 30–32
 - embedded designs, 190, 191–193
 - repeated-measures approach, 58
- Experimenter attention, threat to construct validity, 10*t*
- Experimenter expectancies, threat to construct validity, 10*t*
- Expert sampling, 20
- Explanation, 125–126
- Explanatory design, 126, 129, 161, 164–165
 - statistical tools, 127–128
- Explanatory-sequential approach, 196–197
 - follow-up explanations design, 197, 198–199
 - instrument-development design, 202, 203–204
 - participant-selection design, 197, 199–200
 - theory-development design, 202, 204–205
 - treatment-development design, 202–203, 205–206
- Exploratory-sequential approach, 201–202
- Ex post facto approaches, 120–121, 252
 - action research approach, 216
 - data analysis example, 315–320
 - examples, 122–123
 - internal validity, 120, 121
 - matched grouping, 16
 - posttest designs, 121–122
 - two-group, 121
- External validity, 8
 - qualitative methods and, 4, 142
 - sampling and, 31
 - survey research, 133, 134
 - switching-replications design, 59
 - threats to, 8, 9*t*, 134
- Face validity, 241
- Factorial designs, 75–77
 - mixed designs, 76, 89–91
 - posttest designs, 79–80, 82–85
 - pretest and posttest designs, 77–78, 81–82
 - reference resources, 77
 - Solomon *N*-group design, 93
 - threats to internal validity, 76–77
- Factor notation, 15
- Field research, 32
- Figures, 246–249
- Follow-up explanations design, 197, 198–199
- 4-factor pretest and posttest design, 42–43
- Frequency statistics, 242
- Full transcription, 323
- G*Power, 220, 256
- Gain score analysis (GSA), 253, 262–263
- “Garbage in, garbage out,” 31–32
- Generalization, 133
 - case-to-case, 142
 - construct validity and, 9
 - external validity and, 8, 134
 - survey research and, 133, 134
- Generalized causal inference, 31
- Generic qualitative design, 142–143
- Glass’s delta (Δ), 238
- Gonzales, N., 47
- Graeco-Latin-square design, 63
 - 3-factor design, 227
- Graphing single-case designs, 105
- Graphs, charts, and figures, 246–249
- Grounded theory approach, 145
 - constructivist design, 146, 150–151
 - emerging design, 146, 148*f*, 149
 - reference resources, 147
 - systematic design, 146, 147–149
 - when to use, 146
- Group assignment, 14, 31
 - techniques, 15–17
 - See also* Random assignment
- Health or medical sciences research, 32
- Hedges’s *g*, 238, 266, 318
- Hermeneutic design, 169, 173–174
- Heterogeneity sampling, 20
- Hierarchical linear model (HLM), 254
- Hierarchical linear regression, 128
- Historical control group, 18, 33, 45
 - less common designs, 224
- History, threat to internal validity, 7*t*
- Holistic designs, 144, 229–230
- Homogeneity of variance assumption, 253, 256
- Hypotheses, 22, 30
- Inclusion and control, 14
- Independence assumption, 253, 254
- Independent samples *t* test, 253, 262, 263–267, 293, 297–298, 316–319
- Independent variables (IVs), 5
 - decision tree for statistical tests, 233–235
 - manipulation, 13
 - manipulation check, 235–237
 - predictor variable for nonexperimental research, 5, 127, 128
 - quasi-independent variables, 121

- Inductive reasoning, 142, 218
- Instrumentation threat to internal validity, 7*t*
- Instrument-development design, 202, 203–204
- Intentionality, 169
- Interaction effects, 75
- Inter-item reliability, 239–240
- Internal validity, 7
 - ex post facto designs, 120, 121
 - nonexperimental research and, 4–5, 118
 - regression-discontinuity approach, 52
 - strength of experimental research, 31
 - two-group posttest control group design, 45
- Internal validity, threats to, 7–8*t*, 31
 - factorial designs, 76–77
 - interrupted time-series, 61
 - k*-factor designs, 36–37
 - Latin-square design, 63
 - posttest-only designs, 46
 - sequencing effects, 58
 - single-case approaches, 104
 - Solomon *N*-group design, 93, 94
 - strength of experimental research and, 31
 - within-subjects (repeated measures) approach, 57
- Interrater reliability, 240
- Interrupted time-series (ITS), 58, 61, 69–71
- Interview transcription, 322

- k*-factor designs, 36–37, 61
 - factorial designs, 75

- Latin-square design, 61–64, 72–73
 - Graeco-Latin-square, 227
 - 3-factor Graeco-Latin square, 227
 - 3 × 3, 252, 284–292
 - 3 × 3 repeated, 227–228
- Linearity assumption, 253, 306
- Linear regression, 311–314
- Listening skills, 160
- Logistic regression, 128
- Longitudinal data structure, 58
- Longitudinal studies, 58
- Longitudinal survey design, 135, 136–138

- Manipulation and control, 13
- Manipulation check, 235–237
- Matched grouping, 16, 120
- Matched pairs, 16
- Matching-based assignment, 16, 33
- Maturation, threat to internal validity, 7*t*
- Mean, 242
- Measurement, 30
- Measurement reliability, threats to statistical conclusion validity, 12*t*, 133
- Measures of central tendency, 242–243

- Mediation, 128
 - threat to external validity, 9*t*
- Memoing, 145, 323–325
- Meta-analysis, 221
- Method
 - appropriate designs for research questions, 23–27
 - qualitative research terminology, xviii
 - quantitative research terminology, xvii
- Microsoft Excel, 220
- Missing data, 232
- Mixed-factorial design, 76, 89–91
- Mixed methods, 177–179, 218, 231
 - A-B-A designs, 209–210
 - action research and, 213
 - emergent design, 178
 - multiphase design and program evaluation, 219–220
 - notation, 179
 - quasi-mixed (monostrand conversion), 231*n*
 - reference resources, 180
 - single-case approaches, 208–211
 - See also* Convergent-parallel approach; Embedded approach; Explanatory-sequential approach; Exploratory-sequential approach
- Mixed-subjects approach, 59, 76, 89, 269
- Mono-method bias, 10*t*
- Mono-operation bias, 10*t*
- Multilevel mixed methods design, 181, 187–188
- Multinomial logistic regression, 128
- Multiple baseline design, 105, 111–114
- Multiple comparisons, threats to statistical conclusion validity, 12*t*
- Multiple-group approach, 35. *See also* Between-subjects approaches
- Multiple regression, 128
- Multiple-treatment design, 36, 105
- Multistage models, 126
- Multistage sampling, 20
- Multivariate analysis of variance (MANOVA), 254
- Multivariate approach, 35

- Narrative logic data presentation, 331
- Narrative perspective, 160–161
 - critical design, 161–162, 165*f*, 166
 - data analysis, 322
 - descriptive design, 161, 162–164
 - embedded designs, 189
 - explanatory design, 161, 164–165
 - reference resources, 162
 - when to use, 162
- Natural data presentation, 330
- Naturalistic approach, 142
- Nested designs or models, 19, 189. *See also* Embedded approach

- Nonexperimental research, 117–119
 cause and effect relationships and, 118, 126
 control approaches, 118, 125
 ex post facto designs, 120–123
 independent and dependent variables, 5–6
 predictor and outcome variables, 5, 127
 quasi-experimental designs, 32–34. *See also* Quasi-experimental research
 repeated-measures approach, 58
 statistical tools, 127–129
 validity considerations, 4–5
See also Observational approach; Qualitative methods; Survey approach; specific approaches
- Nonprobability sampling techniques, 20
- Nonrandom assignment, 15
- Nonrandomly assigned control group, 18
- Nonresponse bias, 135
- Normal distribution, 235
- Normality assumptions, 253, 254
- Note-taking, 322. *See also* Memoing
- Novelty effects, threat to construct validity, 10*t*
- Null hypothesis, 22
- Observational approach, 125–126
 embedded-correlational design, 190
 explanatory design, 126, 129
 predictive design, 126–127, 130–132
 statistical tools, 127–129
- Observation notation (O), 15
- “Omnipotent” researcher, 153
- One-factor pretest and posttest control group design, 35–36
- One-group double pretest and posttest design, 225
- One-group pretest and posttest design, 64–65
- One-group repeated-treatment design, 226
- One-group treatment-removed design, 225
- One-way analysis of variance (ANOVA), 253
 between-subjects example, 285–292
 within-subjects example, 278–282
- Online resources, xxiii–xxiv
- Ontological phenomenology, 169
- Open coding, 148, 326
- Operational definition, 6
- Order of treatment, 57, 58. *See also* Sequencing effects
- Outcome variable, 5, 127, 129. *See also* Dependent variables
- Outcome variations, threat to external validity, 9*t*
- Paired samples *t* test, 293, 301–302
- Panel-based data collection, 135
- Parallel-databases design, 181, 183–184
- Parallel-group approach, 32
- Parallel randomized trials, checklist and guidelines, 334–337
- Participant-selection design, 197, 198*f*, 199–200
- Participatory action research (PAR), 213
- Pattern matching designs, 222
- Pearson's *r*, 238, 243–244
- Percentages, 242
- Perspective
 appropriate designs for research questions, 24, 27
 qualitative research terminology, xviii
- Phenomenological perspective, 168
 case study design, 169–170, 174–175
 existential design, 168–169, 170–171
 hermeneutic design, 169, 173–174
 reference resources, 170
 transcendental design, 169, 172–173
 when to use, 170
- Phi coefficient (ϕ), 239
- Point biserial correlation coefficient, 239, 264, 298–299, 317, 319
- Population, 19
- Postpositivism, 218
- Posttest design with historical control for action research, 215
- Posttest-only designs, 33, 43–49, 120–121
 action research approach, 216
 control group design, 45
 ex post facto designs, 121–122
 factorial designs, 79–80, 82–85
 2-factor, 46–48
 3-factor, 48–49
 with historical control group, 44, 224
- Power. *See* Statistical power
- Prediction, 125–126
- Predictive design, 126–127, 130–132
 statistical tools, 128–129
- Predictor variable, 5, 127, 128. *See also* Independent variables
- Pretest and posttest designs, 35–36, 251
 action research approach, 216
 control group plus historical control group design, 224
 data analysis, 252–253
 data analysis, control group design, 255–261
 data analysis, 2-factor design, 261–268
 double design, 223
 examples, 37–43
 factorial designs, 77–78, 81–82
 4-factor, 42–43
 historical control group with pretest and posttest design, 224
 less common designs, 222
 1-factor pretest and posttest control group, 35–36
 one-group double pretest and posttest, 225
 one-group repeated-treatment, 226
 pretest and double-posttest, 65–66
 proxy design, 223
 regression-discontinuity approach, 51, 53–55

- 2-factor, 38–41, 251, 261–268
- 2-factor double-posttest design, 252, 268–278
- 3-factor, 41
 - within-subjects approach one-group design, 64–65
- Probability sampling techniques, 20, 133, 134
- Program evaluation, 213, 219–220
- Propensity-scoring methods, 118
- Proxy pretest-posttest design, 223
- Proxy pretest variables, 33
- Purpose statement, 21
- Purposive sampling, 20

- Qualitative comparative analysis (QCA), 142
- Qualitative data analysis, 220, 321
 - coding, 146, 148, 325–327
 - constant comparative method, 150
 - data preparation/transcription, 322–323
 - data presentation, 330–331
 - memoing, 145, 323–325
 - presenting findings, 327–330
 - programs, 144
- Qualitative methods, xiv, 141–143
 - appropriate designs for research questions, 26–27
 - appropriate research questions, 24
 - external validity and, 142
 - generic design, 142–143
 - terminology, xviii
 - types of research questions, 22–23
 - validity concepts and, 4, 324
 - See also* Case study designs
- Quantitative-informed data presentation, 330
- Quantitative meta-analysis, 221
- Quantitative methods, xiv–xv, 29–30, 218
 - appropriate designs for research questions, 23–26
 - appropriate research questions, 23
 - experimental research, 30–32
 - flowchart, 29f
 - nonexperimental research and, 117–119. *See also* Nonexperimental research terminology, xvii–xviii
 - types of research questions, 22
- Quantitative-qualitative mixed methods. *See* Mixed methods
- Quasi-experimental research, 32–34
 - data analysis example, 305–315
 - less common designs, 222
 - switching-replications design, 60
- Quasi-independent variables, 121

- Random assignment, 15–16, 31
 - posttest-only designs and, 45
 - switching-replications design and, 59
 - wait-list continuation design, 59, 60f
- Randomized control trials (RCTs), 32
- Randomized order of treatment, 57, 58
- Randomized parallel trials, checklist and guidelines, 334–337
- Random sampling, 20
- Random selection, 31, 133
- Range restriction, 12f
- Reactivity to assessment, 10f, 134
- Realist design, 152–153, 154–156
- Regression analysis, 125
 - explanatory design, 126
 - independent variable in, 5
 - predictive design, 126–127
 - statistical tools for predictive designs, 128–129
 - See also* Observational approach
- Regression-discontinuity (RD) approach, 51–56, 118, 252, 305–315
- Regression point-displacement (RPD) designs, 219, 225
- Relational research questions, 22, 25
- Reliability, 239–240
 - threats to statistical conclusion validity, 12f, 133
- Repeated-measures analysis of variance (RM-ANOVA), 128, 254, 269–282
- Repeated-measures approach, 57–59
 - assumptions for data analysis, 254
 - crossover design, 36, 59, 61–64, 68–70
 - data analysis, 254
 - interrupted time-series, 58, 61, 69–71
 - Latin-square design, 61–64, 72–73
 - longitudinal studies, 58
 - pretest and double-posttest design, 65–66
 - reference resources, 64
 - sequencing effects, 7
 - single-case approaches, 58
 - switching-replications design, 59–61, 66–68
 - 3-factor crossover design, 226
 - See also* Within-subjects approach
- Replication, 32
- Reporting data, 32, 330–331. *See also* Research reports
- Reproducibility, 32
- Research
 - appropriate designs for research questions, 23–28
 - qualitative research terminology, xviii
 - quantitative research terminology, xvii
- Research aims, 21
- Research design, xv
 - appropriate designs for research questions, 23–28
 - custom modifications, 221
 - design logic, 13–14. *See also* Control design notations, 15–17
 - discussing in research reports, 221
 - purpose of, 1–2
 - statistical methods and, xvi
 - terminology, xvii–xvix
- Researcher-participant collaboration, 160

- Research questions, 21, 30
 - choosing appropriate designs, 23–28
 - subquestions, 23
 - types of inquiry, 21–23
- Research reports
 - discussing research design, 221
 - parallel randomized trials, 334–337
 - preliminary statistical findings, 231–250. *See also* Statistical findings, reporting
 - qualitative findings presentation, 327–331
 - reporting data, 32
- Resentful demoralization, 11*t*
- Response rate, 134–135
- Sample-related threats to external validity, 9*t*
- Sampling bias, 8
- Sampling strategies, 18–21
 - external validity and, 31, 133, 134
 - nonprobability techniques, 20
 - probability techniques, 20, 133
 - survey research, 133, 134
- Scatterplot, 311–313
- Scientific method, 2–3, 30
- Scott's pi, 240
- Selection bias, 7*t*, 120
- Selective coding, 148
- Semi transcription, 323
- Sensitization, 11*t*
- Sequencing effects
 - counterbalancing, 17, 57, 58, 254
 - threats to internal validity, 7, 8*t*
- Sequential case study single-case design, 210–211
- Simple random sampling, 20
- Single-case approaches, 103–106, 230
 - A-B-A-B-C-B-C design, 108–109
 - A-B-A-B design, 107–108
 - A-B-A design, 104, 106–107
 - A-B-A mixed methods, 209–210
 - changing criterion design, 105, 109–110
 - design notations, 104*t*
 - graphing, 105
 - mixed methods, 208–211
 - multiple baseline design, 111–114
 - repeated measures, 58
 - threats to internal validity, 104
 - See also* Case study designs
- Single-factor randomized group design, 36
- Single-stage models, 126
- Single-subject design, 103
- Small-*n* designs, 103
- Snowball sampling, 20
- Social desirability, 133, 135
- Solomon *N*-group design, 93–94
 - eight-group, 98–101
 - four-group, 94–96
 - six-group, 96–98
 - threats to internal validity, 94
- Spearman's correlation statistic, 238–239, 243
- Special treatment, threat to internal validity, 8*t*
- Sphericity assumption, 254
- SPSS, 220
 - ANOVA independent variable, 5
 - pretest and posttest designs (ANCOVA), 256–259
 - pretest and posttest designs (independent samples *t* test), 263–267
 - pretest and posttest designs (RM ANOVA), 270–277
 - regression-discontinuity approach (ANCOVA/linear regression), 307–314
 - 3-factor crossover design (RM ANOVA), 278–282
 - 3 x 3 Latin square design (ANOVA), 286–290
 - wait-list continuation design (*t* tests), 295–303
- Standard deviation, 242
- Statistical analysis
 - control and, 14
 - data analytics, 220
 - decision tree for choosing tests, 233–235
 - design and, xvi, 251–252
 - ex post facto design, 315–320
 - nonexperimental control approaches, 118, 125
 - observational (nonexperimental) research, 127–129
 - pretest and posttest designs, 252–253
 - pretest and posttest designs, control group, 255–261
 - pretest and posttest designs, double-posttest, 268–278
 - pretest and posttest designs, 2-factor, 261–268
 - regression-discontinuity approach, 305–315
 - 3-factor crossover design, 278–284
 - 3 x 3 Latin square design, 284–292
 - wait-list continuation design, 292–305
 - within-subjects approach (repeated measures), 254
- Statistical conclusion validity, 9, 11, 13
 - checklist, 333
 - survey research and, 133, 134
 - threats to, 11–12*t*
- Statistical findings, reporting, 231
 - chart of symbols, 249–250
 - decision tree for choosing tests, 233–235
 - descriptive statistics, 242–244
 - effect size, 237–239
 - figures, 246–250
 - manipulation check, 235–237
 - normal distribution, 235
 - power, 241–242
 - reliability, 239–240
 - tables, 244–246
 - validation, 240–241
- Statistical power, 220
 - reporting, 241–242
 - threats to statistical conclusion validity, 11*t*
- Statistical regression, threat to internal validity, 7*t*



- Statistical software packages, 220
- Stimulus characteristics, threat to external validity, 9*t*
- Storytelling, 160, 331. *See also* Narrative perspective
- Stratified random assignment, 15, 17*f*
- Stratified sampling, 20
- Subject heterogeneity effects, 12*t*
- Subquestions, 23
- Survey approach, 133
 - cross-sectional design, 135, 136–137
 - longitudinal design, 135, 136–138
 - response rate, 134–135
 - threats to validity, 134–135
- Survey data, 322
- Switching-replications design, 59–61, 66–68
- Systematic design, 146, 147–149
- Systematic sampling, 20
- Tables, 244–246
- Terminology, xvii–xvix
- Testing threat to internal validity, 7*t*
- Test-retest reliability, 239–240
- Test sensitization, 11*t*
- Test validity, 4
- Theoretical sampling, 20
- Theory-development design, 202, 204–205
- Theory-focused qualitative methods, 142
- Theory-guided data presentation, 330
- Theory-guided sampling, 20
- Threats to validity, 4–5. *See also specific types of validity*
- 3-factor crossover design, 226, 252, 278–284
- 3-factor posttest design, 48–49
- 3-factor pretest and posttest design, 41
- 3 x 3 Graeco-Latin square design, 227
- 3 x 3 Latin square design, 252, 284–292
- 3 x 3 repeated Latin-square design, 227–228
- Time-series, 58
 - interrupted, 58, 61, 69–71
 - pretest and posttest designs, 36
- Timing of measurement threat to construct validity, 11*t*
- Transcendental design, 169, 172–173
- Transcription, 322–323
- Transferability, 324
- Treatment-development design, 202–203, 205–206
- Treatment notation, 15
- Treatment-removed design, 225
- Treatment-sensitive factorial structure, 10*t*
- Treatment variations, threat to external validity, 9*t*
- Trend analysis, 127
- Trend-based data collection, 135
- Triangulation design, 142
 - concurrent design, 181, 184*n*
- Trustworthiness of data, 4
- t* test, 22, 127
 - independent samples, 253, 262, 263–267, 293, 297–298, 316–319
 - paired samples (dependent samples), 293, 301–302
- 2-factor crossover design, 68–70
- 2-factor posttest control group design, 46–47
- 2-factor pretest and double posttest designs, 252, 268–278
- 2-factor pretest and posttest designs, 38–41, 251
 - data analysis examples, 261–278
- 2 x 2 factorial posttest design, 79–80, 82–84
- 2 x 2 factorial pretest and posttest design, 77–78, 81–82
- 2 x 2 x 2 factorial posttest design, 85–86
- 2 x 2 x 2 mixed-factorial design, 89–91
- 2 x 3 x 2 factorial posttest design, 87–88
- Two-group ex post facto design, 121
- Two-way analysis of variance (ANOVA), 128, 253
- Uncontrolled case studies, 143
- Unit of analysis, 19
- Univariate comparison of means, 127–128
- Validation, 240–241
- Validity, 1, 3–5
 - checklist, 332–333
 - control and, 13
 - experimental research and, 31–32
 - nonexperimental research and, 118
 - qualitative methods and, 4, 324
 - reporting, 240–241
 - test, 4
 - threats to, 4–5. *See also specific types of validity**See also* Construct validity; External validity; Internal validity; Statistical conclusion validity
- Wait-list continuation design, 59, 60*f*, 252, 292–305
- Wait-list control group, 18, 59
- Washout period, 61, 103
- Website resource, xxiii–xxiv
- Within-subjects approach, 57–58
 - counterbalancing, 57, 58, 254
 - data analysis, 254
 - data analysis (3-factor crossover design), 278–284
 - factorial designs, 76, 81–82
 - Latin-square design, 61–64, 72–73
 - one-group double pretest and posttest design, 225
 - one-group pretest and posttest design, 64–65
 - switching-replications design, 59–61, 66–68
 - threats to internal validity, 57*See also* Repeated-measures approach