# Brief Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables, Figures, and Diagrams</td>
<td>xi</td>
</tr>
<tr>
<td>Foreword</td>
<td>xiv</td>
</tr>
<tr>
<td>Preface</td>
<td>xvii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>xxv</td>
</tr>
<tr>
<td>About the Authors</td>
<td>xxvi</td>
</tr>
<tr>
<td>1. A Primer of the Scientific Method and Relevant Components</td>
<td>1</td>
</tr>
<tr>
<td>PART I. QUANTITATIVE METHODS FOR EXPERIMENTAL AND QUASI-EXPERIMENTAL RESEARCH</td>
<td>29</td>
</tr>
<tr>
<td>2. Between-Subjects Approach</td>
<td>35</td>
</tr>
<tr>
<td>3. Regression-Discontinuity Approach</td>
<td>51</td>
</tr>
<tr>
<td>4. Within-Subjects Approach</td>
<td>57</td>
</tr>
<tr>
<td>5. Factorial Designs</td>
<td>75</td>
</tr>
<tr>
<td>6. Solomon $N$-Group Designs</td>
<td>93</td>
</tr>
<tr>
<td>7. Single-Case Approach</td>
<td>103</td>
</tr>
<tr>
<td>PART II. QUANTITATIVE METHODS FOR NONEXPERIMENTAL RESEARCH</td>
<td>117</td>
</tr>
<tr>
<td>8. Ex Post Facto and Posttest-Only Designs</td>
<td>120</td>
</tr>
<tr>
<td>9. Observational Approach</td>
<td>125</td>
</tr>
<tr>
<td>10. Survey Approach</td>
<td>133</td>
</tr>
<tr>
<td>PART III. QUALITATIVE METHODS</td>
<td>141</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>11. Grounded Theory Perspective</td>
<td>145</td>
</tr>
<tr>
<td>12. Ethnographic Perspective</td>
<td>152</td>
</tr>
<tr>
<td>13. Narrative Perspective</td>
<td>160</td>
</tr>
<tr>
<td>14. Phenomenological Perspective</td>
<td>168</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PART IV. MIXED METHODS</th>
<th>177</th>
</tr>
</thead>
<tbody>
<tr>
<td>15. Convergent-Parallel Approach</td>
<td>181</td>
</tr>
<tr>
<td>16. Embedded Approach</td>
<td>189</td>
</tr>
<tr>
<td>17. Explanatory-Sequential Approach</td>
<td>196</td>
</tr>
<tr>
<td>18. Exploratory-Sequential Approach</td>
<td>201</td>
</tr>
<tr>
<td>19. Mixed Methods, Case Studies, and Single-Case Approaches</td>
<td>208</td>
</tr>
<tr>
<td>20. Action Research Approaches</td>
<td>212</td>
</tr>
<tr>
<td>21. Conclusion</td>
<td>218</td>
</tr>
</tbody>
</table>

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
1. A Primer of the Scientific Method and Relevant Components
   The Scientific Method
   Validity and Research Designs
   Independent and Dependent Variables
   Internal Validity
   External Validity
   Construct Validity
   Statistical Conclusion Validity
   Design Logic
   Control
   Design Notations
   Assignment Techniques
   Comparison and Control Groups
   Sampling Strategies
   Probability Sampling Techniques
   Nonprobability Sampling Techniques
PART I. QUANTITATIVE METHODS FOR EXPERIMENTAL AND QUASI-EXPERIMENTAL RESEARCH

Experimental Research
Quasi-Experimental Research

2. Between-Subjects Approach
Pretest and Posttest Designs
\textit{k-Factor Designs}
Posttest Designs

3. Regression-Discontinuity Approach

4. Within-Subjects Approach
Repeated-Measures Approach
\textit{Switching-Replications Design: A Primer}
\textit{Crossover and Latin-Square Designs}

5. Factorial Designs

6. Solomon \textit{N}-Group Design

7. Single-Case Approach

PART II. QUANTITATIVE METHODS FOR NONEXPERIMENTAL RESEARCH

8. Ex Post Facto and Posttest-Only Designs

9. Observational Approach
Explanatory Design
Predictive Design
\textit{Statistical Tools}
\textit{Predicting a Single Variable for the Predictive Design}
\textit{Predicting More Than One Variable}

10. Survey Approach
Primary Threats to Validity
Cross-Sectional Design
Longitudinal Design
PART III. QUALITATIVE METHODS

The Case Study

11. Grounded Theory Perspective
   Systematic Design
   Emerging Design
   Constructivist Design

12. Ethnographic Perspective
   Realist Design
   Critical Design
   Case Study Design

13. Narrative Perspective
   Descriptive Design
   Explanatory Design
   Critical Design

14. Phenomenological Perspective
   Existential Design
   Transcendental Design
   Hermeneutic Design
   Case Study Design

PART IV. MIXED METHODS

Mixed Methods Legend

15. Convergent-Parallel Approach
   Parallel-Databases Design
   Data-Transformation Design
   Data-Validation Design
   Multilevel Design

16. Embedded Approach
   Embedded-Experiment Design
   Embedded-Correlational Design
   Embedded Case Study Design

17. Explanatory-Sequential Approach
   Follow-Up Explanations Design
   Participant-Selection Design
18. Exploratory-Sequential Approach
   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
   Sequential Case Study Single-Case Design 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 1.2. Threats to External Validity 9
Table 1.3. Threats to Construct Validity 10
Table 1.4. Threats to Statistical Conclusion Validity 11
Table 1.5. Control Techniques for Experimental, Quasi-Experimental, and Nonexperimental Research 14
Table 1.6. Design Notations 15
Table 1.7. Group Assignment Design Notations 15
Table 7.1. Design Notations for Single-Case Approaches 104
Table 20.1. Design Notations for the Action Research Approach 215

FIGURES ♦

Figure 1.1. The Scientific Method 3
Figure 1.2. Conceptual and Operational Definitions 6
Figure 1.3. Example of a Stratified Random-Assignment Technique 17
Figure 1.4. Example of a Sample Extracted From a Population 19
Figure I.1. Quantitative Method Flowchart 29
Figure I.2. Double Pretest Design for Quasi-Experimental Research 33
Figure 3.1. Sample of a Cutoff Score 52
Figure 4.1. Wait-List Continuation Design (for Random Assignment) 60
Figure 4.2. Switching-Replications Design (for Nonrandom Assignment) 60
Figure 11.1. Systematic Design 147
Figure 11.2. Emerging Design 148
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 Design 53
Diagram 3.2. Regression-Discontinuity Pretest–Posttest Control Group Design 54
Diagram 4.1. A Repeated-Measures Approach 3-Factor Crossover Design 62
Diagram 4.2. Example of a 3 x 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 × 2 × 2 Factorial Posttest Design (Between-Subjects) 85
Diagram 5.6. 2 × 3 × 2 Factorial Posttest Design (Between-Subjects) 87
Diagram 5.7. 2 × 2 × Mixed Factorial Design (Mixed-Subjects) 89
Diagram 6.1. Solomon Four-Group Design 94
Diagram 6.2. Solomon Six-Group Design 96
Diagram 7.1. A-B-A Design 106
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 Design 130
Diagram 9.3. Predictive Design 2 131
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
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 investiga-
tion 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 investiga-
tion, so students must develop expertise in all three. Certainly one must
have sound measurement to be able to draw conclusions about the under-
lying 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 provid-
ing 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 cov-
ers 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 recom-
mended 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
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

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD\textsubscript{1}</td>
<td>The <em>method</em> provides the theoretical, philosophical, and data-analytic stance (e.g., a quantitative method\textsubscript{1}).</td>
</tr>
<tr>
<td>RESEARCH\textsubscript{2}</td>
<td><em>Research</em> 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\textsubscript{1} and experimental research\textsubscript{2}).</td>
</tr>
</tbody>
</table>

(Continued)
### QUANTITATIVE

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPROACH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>The <em>approach</em> 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&lt;sub&gt;1&lt;/sub&gt;, experimental research&lt;sub&gt;2&lt;/sub&gt; with a between-subjects approach&lt;sub&gt;3&lt;/sub&gt;).</td>
</tr>
<tr>
<td>DESIGN&lt;sub&gt;4&lt;/sub&gt;</td>
<td>The <em>design</em> 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&lt;sub&gt;1&lt;/sub&gt;, experimental research&lt;sub&gt;2&lt;/sub&gt; with a between-subjects approach&lt;sub&gt;3&lt;/sub&gt; and a pre- and posttest control group design&lt;sub&gt;4&lt;/sub&gt;).</td>
</tr>
</tbody>
</table>

### QUALITATIVE

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD&lt;sub&gt;1&lt;/sub&gt;</td>
<td>The <em>method</em> provides the theoretical, philosophical, and data analytic stance (e.g., a qualitative method&lt;sub&gt;1&lt;/sub&gt;).</td>
</tr>
<tr>
<td>RESEARCH&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Research for the qualitative method is nonexperimental (e.g., a qualitative method&lt;sub&gt;1&lt;/sub&gt; and nonexperimental research&lt;sub&gt;2&lt;/sub&gt;).</td>
</tr>
<tr>
<td>PERSPECTIVE&lt;sub&gt;3&lt;/sub&gt;</td>
<td>The <em>perspective</em> 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&lt;sub&gt;1&lt;/sub&gt;, nonexperimental research&lt;sub&gt;2&lt;/sub&gt; with an ethnographic perspective&lt;sub&gt;3&lt;/sub&gt;).</td>
</tr>
<tr>
<td>DESIGN&lt;sub&gt;4&lt;/sub&gt;</td>
<td>The <em>design</em> 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&lt;sub&gt;1&lt;/sub&gt;, nonexperimental research&lt;sub&gt;2&lt;/sub&gt; with an ethnographic&lt;sub&gt;3&lt;/sub&gt; and a case study design&lt;sub&gt;4&lt;/sub&gt;).</td>
</tr>
</tbody>
</table>

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
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.

![The Scientific Method](image)

**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 “breakdown” 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

![Diagram showing the relationship between variables, conceptual definitions, and operational definitions.](image-url)
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>
<thead>
<tr>
<th>Threat</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>History</td>
<td>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)</td>
</tr>
<tr>
<td>Maturation</td>
<td>The natural process of changing, growing, and learning over time</td>
</tr>
<tr>
<td>Testing</td>
<td>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)</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>The change in a measuring instrument over time (i.e., some instruments undergo revisions)</td>
</tr>
<tr>
<td>Statistical regression</td>
<td>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)</td>
</tr>
<tr>
<td>Selection bias</td>
<td>Also known as <em>selection effect</em>; 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</td>
</tr>
</tbody>
</table>

(Continued)
Threat | Explanation
---|---
Attrition | The loss of participants during the term of the experiment (also known as drop-out or subject mortality)
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 | Related to within-subject (repeated-measures) approaches and also known as multiple-treatment interference, fatigue effects, and practice effects; can be separated into order effects (i.e., the order in which participants receive the treatment can affect the results) and carryover effects (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.
Chapter 1  A Primer of the Scientific Method and Relevant Components

Table 1.2  Threats to External Validity

<table>
<thead>
<tr>
<th>Threat</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample characteristics</td>
<td>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)</td>
</tr>
<tr>
<td>Stimulus characteristics and settings</td>
<td>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)</td>
</tr>
<tr>
<td>Treatment variations</td>
<td>Variations in the same treatment or the combination of multiple or partial treatments that account for different results</td>
</tr>
<tr>
<td>Outcome variations</td>
<td>Observing the effect of one type of outcome differs when alternate outcomes are observed</td>
</tr>
<tr>
<td>Context-dependent mediation</td>
<td>Mediating variables related to outcomes differ between contexts or settings</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Threat</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attention and contact with</strong></td>
<td>Similar to special treatment; 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)</td>
</tr>
<tr>
<td>participants**</td>
<td></td>
</tr>
<tr>
<td><strong>Single operations and narrow</strong></td>
<td>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)</td>
</tr>
<tr>
<td><strong>stimulus sampling</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Experimenter expectancies</strong></td>
<td>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)</td>
</tr>
<tr>
<td><strong>Cues of the experimental</strong></td>
<td>Sources of influence conveyed to prospective participants (e.g., rumors, information passed along from previous participants)</td>
</tr>
<tr>
<td><strong>situation</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Novelty effects</strong></td>
<td>The novelty of being in a new or innovative context</td>
</tr>
<tr>
<td><strong>Inadequate explication of</strong></td>
<td>The construct under investigation is not appropriately defined conceptually, leading to inadequate measurement (i.e., operationalization)</td>
</tr>
<tr>
<td><strong>constructs</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Construct confounding</strong></td>
<td>Multiple constructs not clearly identified and accounted for operationally</td>
</tr>
<tr>
<td><strong>Mono-operation bias</strong></td>
<td>An operationalization (i.e., measurement) does not appropriately represent the construct under investigation, leading to measuring unintended constructs</td>
</tr>
<tr>
<td><strong>Mono-method bias</strong></td>
<td>All measurement techniques are the same as a means to measure the construct under investigation</td>
</tr>
<tr>
<td><strong>Confounding constructs with</strong></td>
<td>All the levels of a construct are not fully accounted for through the appropriate measurement and reporting tools</td>
</tr>
<tr>
<td><strong>levels of constructs</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Treatment sensitive factorial</strong></td>
<td>The interpretation and structure of a measure change as a result of the treatment</td>
</tr>
<tr>
<td><strong>structure</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Reactivity to assessment</strong></td>
<td>The participants' awareness of being studied may influence the outcome; also known as acquiescence bias, social desirability, and the Hawthorne or observer effect; 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)</td>
</tr>
</tbody>
</table>
Threatsen sensitization

Also known as *pretest sensitization*; 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

<table>
<thead>
<tr>
<th>Threat</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low statistical power</td>
<td>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.</td>
</tr>
<tr>
<td>Assumption violation of statistical tests</td>
<td>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.</td>
</tr>
</tbody>
</table>

(Continued)
### Table 1.4 (Continued)

<table>
<thead>
<tr>
<th>Threat</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Error rate problem</strong></td>
<td>Statistical significance can be artificially inflated when performing multiple pairwise tests; it is also referred to as family-wise error rate (i.e., the probability of making a Type I error when performing multiple pairwise analyses).</td>
</tr>
<tr>
<td><strong>Restriction of range</strong></td>
<td>A lack of variability between variables weakens the relationship and lowers statistical power.</td>
</tr>
<tr>
<td><strong>Extraneous variance in the experimental setting</strong></td>
<td>Variations within the experimental setting (e.g., temperature) may inflate error.</td>
</tr>
<tr>
<td><strong>Inaccurate effect size estimation</strong></td>
<td>Some statistical analyses can overestimate or underestimate the size of an effect.</td>
</tr>
<tr>
<td><strong>Variability in the procedures</strong></td>
<td>Also referred to as unreliability of treatment implementation, 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).</td>
</tr>
<tr>
<td><strong>Subject heterogeneity</strong></td>
<td>The variability of participant demographics (e.g., age, race, ethnicity, background) may create unaccounted-for variations in the findings.</td>
</tr>
<tr>
<td><strong>Unreliability of the measures</strong></td>
<td>Measures maintain certain levels of validity and reliability (pertaining to psychometric principles), and lack of reliability causes inconsistency in measurement.</td>
</tr>
<tr>
<td><strong>Multiple comparisons and error rates</strong></td>
<td>The use of multiple dependent variables across conditions and multiple statistical analyses creates greater opportunities for error variance.</td>
</tr>
</tbody>
</table>

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


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

<table>
<thead>
<tr>
<th>Type of Control</th>
<th>Experimental and Quasi-Experimental Research</th>
<th>Nonexperimental Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manipulation</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Elimination</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Inclusion</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Group or condition assignment</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Statistical procedures</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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 \( O_1, O_2, \ldots O_n \).

**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.

### 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...
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:


**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) | ↓ | ↓ | ↓ |
| 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).
Comparable 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](image)

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 nth 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 *haphazard* 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:


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.

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD\textsubscript{1}</td>
<td>The method provides the theoretical, philosophical, and data analytic stance (e.g., a quantitative method\textsubscript{1}).</td>
</tr>
<tr>
<td>RESEARCH\textsubscript{2}</td>
<td>Research 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\textsubscript{1}, experimental research\textsubscript{2}).</td>
</tr>
<tr>
<td>APPROACH\textsubscript{3}</td>
<td>The approach 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\textsubscript{1}, experimental research\textsubscript{2}, with a between-subjects approach\textsubscript{3}).</td>
</tr>
<tr>
<td>DESIGN\textsubscript{4}</td>
<td>The design 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\textsubscript{1}, experimental research\textsubscript{2}, with a between-subjects approach\textsubscript{3}, and a pre- and posttest control group design\textsubscript{4}).</td>
</tr>
</tbody>
</table>
### QUALITATIVE

<table>
<thead>
<tr>
<th>LEVEL</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD₁</td>
<td>The method provides the theoretical, philosophical, and data analytic stance (e.g., a qualitative method).</td>
</tr>
<tr>
<td>RESEARCH₂</td>
<td>Research for the qualitative method is nonexperimental (e.g., a qualitative method and nonexperimental research).</td>
</tr>
<tr>
<td>PERSPECTIVE₃</td>
<td>The perspective 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, nonexperimental research with an ethnographic perspective).</td>
</tr>
<tr>
<td>DESIGN₄</td>
<td>The design 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, nonexperimental research with an ethnographic and a case study design).</td>
</tr>
</tbody>
</table>

### Type of Research Question

<table>
<thead>
<tr>
<th>Level</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD₁</td>
<td>Quantitative or qualitative</td>
</tr>
<tr>
<td>RESEARCH₂</td>
<td>Experimental, quasi-experimental, or nonexperimental</td>
</tr>
<tr>
<td>APPROACH₃</td>
<td>Quantitative or qualitative methodological variant</td>
</tr>
<tr>
<td>DESIGN₄</td>
<td>Any design variant found under the quantitative or qualitative method</td>
</tr>
</tbody>
</table>

### Example 1.1

**Descriptive**

<table>
<thead>
<tr>
<th>Level</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD₁</td>
<td>Quantitative</td>
</tr>
<tr>
<td>RESEARCH₂</td>
<td>Nonexperimental</td>
</tr>
</tbody>
</table>
**Chapter 1  A Primer of the Scientific Method and Relevant Components**  ♦  25

### Descriptive

<table>
<thead>
<tr>
<th>APPRAOCH 3</th>
<th>Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESIGN 4</td>
<td>Cross-sectional</td>
</tr>
</tbody>
</table>

*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

### Relational

<table>
<thead>
<tr>
<th>Level</th>
<th>To what extent do levels of perceived anxiety predict performance on standardized testing?</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD 1</td>
<td>Quantitative</td>
</tr>
<tr>
<td>RESEARCH 2</td>
<td>Nonexperimental</td>
</tr>
<tr>
<td>APPRAOCH 3</td>
<td>Observational</td>
</tr>
<tr>
<td>DESIGN 4</td>
<td>Predictive</td>
</tr>
</tbody>
</table>

*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

### Comparison

<table>
<thead>
<tr>
<th>Level</th>
<th>How do the groups differ between the high-anxiety and low-anxiety conditions in terms of test performance?</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD 1</td>
<td>Quantitative</td>
</tr>
<tr>
<td>RESEARCH 2</td>
<td>Experimental</td>
</tr>
</tbody>
</table>

*(Continued)*
Example 1.3 (Continued)

<table>
<thead>
<tr>
<th>APPROACH</th>
<th>DESIGN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between-subjects</td>
<td>2-factor posttest</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Level</th>
<th>How do the groups differ when exposed to the high-anxiety and low-anxiety conditions in terms of test performance?</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD₁</td>
<td>Quantitative</td>
</tr>
<tr>
<td>RESEARCH₂</td>
<td>Experimental</td>
</tr>
<tr>
<td>APPROACH₃</td>
<td>Within-subjects (repeated-measures) approach can be used.</td>
</tr>
<tr>
<td>DESIGN₄</td>
<td>2-factor crossover</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Level</th>
<th>Central Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>METHOD₁</td>
<td>What are the experiences of parents who have children diagnosed with a pervasive developmental disorder (PDD)?</td>
</tr>
<tr>
<td>RESEARCH₂</td>
<td>Qualitative</td>
</tr>
</tbody>
</table>
Chapter 1  A Primer of the Scientific Method and Relevant Components ♦ 27

Central Question

<table>
<thead>
<tr>
<th>PERSPECTIVE</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESIGN</td>
<td>Descriptive</td>
</tr>
</tbody>
</table>

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

Central Question

<table>
<thead>
<tr>
<th>Level</th>
<th>Method</th>
<th>Research</th>
<th>Perspective</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Qualitative</td>
<td>Nonexperimental</td>
<td>Ethnographic</td>
<td>Realist</td>
</tr>
</tbody>
</table>

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 statisti­
   cal 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 resources.

<table>
<thead>
<tr>
<th>Method</th>
<th>Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>Experimental and Quasi-Experimental</td>
</tr>
<tr>
<td>Approach</td>
<td>Between-Subjects</td>
</tr>
<tr>
<td></td>
<td>Within-Subjects</td>
</tr>
<tr>
<td></td>
<td>Regression Discontinuity</td>
</tr>
<tr>
<td></td>
<td>Single-Case</td>
</tr>
<tr>
<td>Design</td>
<td>Pretest-Posttest</td>
</tr>
<tr>
<td></td>
<td>Posttest</td>
</tr>
<tr>
<td></td>
<td>Factorial</td>
</tr>
<tr>
<td></td>
<td>Solomon N-Group</td>
</tr>
<tr>
<td></td>
<td>A-B</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest&lt;sub&gt;1&lt;/sub&gt;</th>
<th>Pretest&lt;sub&gt;2&lt;/sub&gt;</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>O&lt;sub&gt;1&lt;/sub&gt;</td>
<td>O&lt;sub&gt;2&lt;/sub&gt;</td>
<td>X</td>
<td>O&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
<tr>
<td>2</td>
<td>NR</td>
<td>O&lt;sub&gt;1&lt;/sub&gt;</td>
<td>O&lt;sub&gt;2&lt;/sub&gt;</td>
<td></td>
<td>O&lt;sub&gt;3&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

**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 O<sub>1</sub> to O<sub>2</sub>, and then from O<sub>2</sub> to O<sub>3</sub>. 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 \(X_A, X_B, \text{ and } X_C\) (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:


### Diagram 2.1 Pretest and Posttest Control Group Design

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( O_1 )</td>
<td>( X )</td>
<td>( O_2 )</td>
</tr>
<tr>
<td>2</td>
<td>( O_1 )</td>
<td>( - )</td>
<td>( O_2 )</td>
</tr>
</tbody>
</table>

**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

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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 (n = 22)</td>
<td>TELD-3, ECBI</td>
<td>Family-centered intervention</td>
<td>TELD-3, ECBI</td>
</tr>
<tr>
<td>R</td>
<td>2 (n = 19)</td>
<td>TELD-3, ECBI</td>
<td>—</td>
<td>TELD-3, ECBI</td>
</tr>
</tbody>
</table>

Diagram 2.2 2-Factor Pretest and Posttest Control Group Design

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$O_1$</td>
<td>$X_A$</td>
<td>$O_2$</td>
</tr>
<tr>
<td>2</td>
<td>$O_1$</td>
<td>$X_B$</td>
<td>$O_2$</td>
</tr>
<tr>
<td>3</td>
<td>$O_1$</td>
<td>—</td>
<td>$O_2$</td>
</tr>
</tbody>
</table>

Time ▶
Example for Diagram 2.2


**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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 (n = 20)</td>
<td>CBCL, SBCL</td>
<td>Problem solving</td>
<td>CBCL, SBCL</td>
</tr>
<tr>
<td>R</td>
<td>2 (n = 19)</td>
<td>CBCL, SBCL</td>
<td>Relationship therapy</td>
<td>CBCL, SBCL</td>
</tr>
<tr>
<td>R</td>
<td>3 (n = 17)</td>
<td>CBCL, SBCL</td>
<td>—</td>
<td>CBCL, SBCL</td>
</tr>
</tbody>
</table>
Example for Diagram 2.3


**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 pretesting 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)
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

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁</td>
<td>Xₐ</td>
<td>O₂</td>
</tr>
<tr>
<td>2</td>
<td>O₁</td>
<td>Xₐ</td>
<td>O₂</td>
</tr>
<tr>
<td>3</td>
<td>O₁</td>
<td>Xₐ</td>
<td>O₂</td>
</tr>
<tr>
<td>4</td>
<td>O₁</td>
<td>Xₐ</td>
<td>O₂</td>
</tr>
</tbody>
</table>

Example for Diagram 2.5


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

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 (n = 15)</td>
<td>Interest</td>
<td>Cognitive interest</td>
<td>Interest, achievement, time on reading</td>
</tr>
<tr>
<td>R</td>
<td>2 (n = 15)</td>
<td>Interest</td>
<td>Emotional interest</td>
<td>Interest, achievement, time on reading</td>
</tr>
<tr>
<td>R</td>
<td>3 (n = 15)</td>
<td>Interest</td>
<td>Text-only</td>
<td>Interest, achievement, time on reading</td>
</tr>
</tbody>
</table>

| 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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 (n = 105)</td>
<td>Graph interpretation and construction, metacognitive questionnaire</td>
<td>COOP+META</td>
<td>Graph interpretation and construction, metacognitive questionnaire</td>
</tr>
<tr>
<td>R</td>
<td>2 (n = 95)</td>
<td>Graph interpretation and construction, metacognitive questionnaire</td>
<td>IND+META</td>
<td>Graph interpretation and construction, metacognitive questionnaire</td>
</tr>
<tr>
<td>R</td>
<td>3 (n = 91)</td>
<td>Graph interpretation and construction, metacognitive questionnaire</td>
<td>COOP</td>
<td>Graph interpretation and construction, metacognitive questionnaire</td>
</tr>
<tr>
<td>R</td>
<td>4 (n = 93)</td>
<td>Graph interpretation and construction, metacognitive questionnaire</td>
<td>IND</td>
<td>Graph interpretation and construction, metacognitive questionnaire</td>
</tr>
</tbody>
</table>

**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: $O_2$), 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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Test</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>$O_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>NR</td>
<td></td>
<td>X</td>
<td>$O_1$</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>O₁</td>
</tr>
<tr>
<td>2</td>
<td>—</td>
<td>O₁</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td></td>
</tr>
</tbody>
</table>

---

**Example for Diagram 2.7**


**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).
### Assignment R Group Treatment Posttest

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 (n = 17)</td>
<td>Computer-mediated, problem-based learning</td>
<td>Learning outcomes, time on task, learning issues</td>
</tr>
<tr>
<td>R</td>
<td>2 (n = 17)</td>
<td>—</td>
<td>Learning outcomes, time on task, learning issues</td>
</tr>
</tbody>
</table>

**Diagram 2.8 2-Factor Posttest Control Group Design**

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_A$</td>
<td>$O_1$</td>
</tr>
<tr>
<td>2</td>
<td>$X_B$</td>
<td>$O_1$</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>$O_1$</td>
</tr>
</tbody>
</table>

**Example for Diagram 2.8**


*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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_s</td>
<td>1 ( n = 94 )</td>
<td>Follow-up with peer interaction</td>
<td>Attitude survey, course completion</td>
</tr>
<tr>
<td>R_s</td>
<td>2 ( n = 98 )</td>
<td>Follow-up without peer interaction</td>
<td>Attitude survey, course completion</td>
</tr>
<tr>
<td>R_s</td>
<td>3 ( n = 88 )</td>
<td>—</td>
<td>Attitude survey, course completion</td>
</tr>
</tbody>
</table>

**Diagram 2.9** 2-Factor Posttest Design

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( X_A )</td>
<td>( O_1 )</td>
</tr>
<tr>
<td>2</td>
<td>( X_B )</td>
<td>( O_1 )</td>
</tr>
</tbody>
</table>

**Example for Diagram 2.9**


**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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 ((n = 12))</td>
<td>Integrated Psychological Treatment with therapy dog</td>
<td>Positive and Negative Symptoms scale, Living Skills Profile, Quality of Life, Satisfaction</td>
</tr>
<tr>
<td>R</td>
<td>2 ((n = 9))</td>
<td>Integrated Psychological Treatment</td>
<td>Positive and Negative Symptoms scale, Living Skills Profile, Quality of Life, Satisfaction</td>
</tr>
</tbody>
</table>

**Diagram 2.10** 3-Factor Posttest Design

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(X_A)</td>
<td>(O_1)</td>
</tr>
<tr>
<td>2</td>
<td>(X_B)</td>
<td>(O_1)</td>
</tr>
<tr>
<td>3</td>
<td>(X_C)</td>
<td>(O_1)</td>
</tr>
</tbody>
</table>

Example for Diagram 2.10

**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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 ( n = 64 )</td>
<td>Group-based high-intensity training</td>
<td>VO2, BMI, QOL</td>
</tr>
<tr>
<td>R</td>
<td>2 ( n = 64 )</td>
<td>Home-based high-intensity training</td>
<td>VO2, BMI, QOL</td>
</tr>
<tr>
<td>R</td>
<td>3 ( n = 64 )</td>
<td>Home-based low-intensity training</td>
<td>VO2, BMI, QOL</td>
</tr>
</tbody>
</table>

**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).
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:


---

**Diagram 3.1** Regression-Discontinuity Pretest–Posttest Control Group Design

<table>
<thead>
<tr>
<th>Pretest</th>
<th>Assignment</th>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_A$</td>
<td>C</td>
<td>1</td>
<td>X</td>
<td>$O_2$</td>
</tr>
<tr>
<td>$O_A$</td>
<td>C</td>
<td>2</td>
<td>—</td>
<td>$O_2$</td>
</tr>
</tbody>
</table>

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

---

**Example for Diagram 3.1**

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


<table>
<thead>
<tr>
<th>Pretest</th>
<th>Assignment</th>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEMI-PM</td>
<td>C</td>
<td>1 (n = 51)</td>
<td>TEKS</td>
<td>TEMI-PM</td>
</tr>
<tr>
<td>TEMI-PM</td>
<td>C</td>
<td>2 (n = 215)</td>
<td>—</td>
<td>TEMI-PM</td>
</tr>
</tbody>
</table>

Note: Group = First and second graders.

Diagram 3.2  Regression-Discontinuity Pretest-Posttest Control Group Design

<table>
<thead>
<tr>
<th>Pretest</th>
<th>Assignment</th>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>O_A</td>
<td>C</td>
<td>1</td>
<td>X</td>
<td>O_2</td>
</tr>
<tr>
<td>O_A</td>
<td>C</td>
<td>2</td>
<td>—</td>
<td>O_2</td>
</tr>
</tbody>
</table>

Note: Time ➤
Example for Diagram 3.2


*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


<table>
<thead>
<tr>
<th>Pretest Assignment Group Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>English placement examination C 1 (n = 94) Remedial English program GPA</td>
<td></td>
</tr>
<tr>
<td>English placement examination C 2 (n = 103) — GPA</td>
<td></td>
</tr>
</tbody>
</table>

**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.
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)

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Test</th>
<th>Treatment</th>
<th>Test</th>
<th>Time Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X_A</td>
<td>O_1</td>
<td>-</td>
<td>O_2</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>O_3</td>
<td>X_A</td>
<td>O_4</td>
<td></td>
</tr>
</tbody>
</table>

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)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O_1</td>
<td>X_A</td>
<td>O_2</td>
<td>-</td>
<td>O_3</td>
</tr>
<tr>
<td>2</td>
<td>O_4</td>
<td>-</td>
<td>O_5</td>
<td>X_A</td>
<td>O_6</td>
</tr>
</tbody>
</table>

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, Stutken, 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

<table>
<thead>
<tr>
<th>Subject</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_A$</td>
<td>$O_1$</td>
<td>$X_B$</td>
<td>$O_2$</td>
<td>$X_C$</td>
<td>$O_3$</td>
</tr>
<tr>
<td>2</td>
<td>$X_B$</td>
<td>$O_1$</td>
<td>$X_C$</td>
<td>$O_2$</td>
<td>$X_A$</td>
<td>$O_3$</td>
</tr>
<tr>
<td>3</td>
<td>$X_C$</td>
<td>$O_1$</td>
<td>$X_A$</td>
<td>$O_2$</td>
<td>$X_B$</td>
<td>$O_3$</td>
</tr>
<tr>
<td>4</td>
<td>$X_C$</td>
<td>$O_1$</td>
<td>$X_B$</td>
<td>$O_2$</td>
<td>$X_A$</td>
<td>$O_3$</td>
</tr>
<tr>
<td>5</td>
<td>$X_A$</td>
<td>$O_1$</td>
<td>$X_C$</td>
<td>$O_2$</td>
<td>$X_B$</td>
<td>$O_3$</td>
</tr>
<tr>
<td>6</td>
<td>$X_B$</td>
<td>$O_1$</td>
<td>$X_A$</td>
<td>$O_2$</td>
<td>$X_C$</td>
<td>$O_3$</td>
</tr>
</tbody>
</table>

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 $3 \times 3$ 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 $3 \times 3$ 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 $3 \times 3$ Latin-Square Design

<table>
<thead>
<tr>
<th>Group</th>
<th>Setting</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ($n = 25$)</td>
<td>Office</td>
<td>Live: A, Recorded: B, Combo: C</td>
</tr>
<tr>
<td>2 ($n = 25$)</td>
<td>Home</td>
<td>Live: B, Recorded: C, Combo: A</td>
</tr>
<tr>
<td>3 ($n = 25$)</td>
<td>Field</td>
<td>Live: C, Recorded: A, Combo: B</td>
</tr>
</tbody>
</table>

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.

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:


**Diagram 4.3** Pretest and Posttest Design (One-Group)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$O_1$</td>
<td>$X$</td>
<td>$O_2$</td>
</tr>
</tbody>
</table>

**Example for Diagram 4.3**


**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)
Table 4.1: Assignment, Group, Pretest, Treatment, and Posttest

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>1 (n = 35)</td>
<td>Asthma Information Quiz, Child Attitude Toward Illness Scale</td>
<td>Okay With Asthma™</td>
<td>Asthma Information Quiz, Child Attitude Toward Illness Scale</td>
</tr>
</tbody>
</table>

Diagram 4.4: Pretest and Multiple-Posttest Design

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest₁</th>
<th>Posttest₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁</td>
<td>X</td>
<td>O₂</td>
<td>O₃</td>
</tr>
<tr>
<td>2</td>
<td>O₁</td>
<td>-</td>
<td>O₂</td>
<td>O₃</td>
</tr>
</tbody>
</table>

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

Example for Diagram 4.4


**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).

### Diagram 4.5 Switching-Replications Design (Experimental)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
<th>Posttest (2-Week Delay)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rs</td>
<td>1 (n = 41)</td>
<td>TV, El, SE, EP</td>
<td>Value-Reappraisal</td>
<td>TV, El, SE, EP</td>
<td>TV, El, SE, EP</td>
</tr>
<tr>
<td>Rs</td>
<td>2 (n = 41)</td>
<td>TV, El, SE, EP</td>
<td>—</td>
<td>TV, El, SE, EP</td>
<td>TV, El, SE, EP</td>
</tr>
</tbody>
</table>

#### Example for Diagram 4.5


**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).

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest (Baseline)</th>
<th>Treatment</th>
<th>Midtest (3 Months)</th>
<th>Treatment</th>
<th>Posttest (6 Months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 ($n = 8$)</td>
<td>Mood, depression</td>
<td>Pleasant-events-focused treatment</td>
<td>Mood, depression</td>
<td>---</td>
<td>Mood, depression</td>
</tr>
<tr>
<td>R</td>
<td>2 (wait-list; $n = 7$)</td>
<td>Mood, depression</td>
<td>---</td>
<td>Mood, depression</td>
<td>Pleasant-events-focused treatment</td>
<td>Mood, depression</td>
</tr>
</tbody>
</table>

Diagram 4.6 Switching-Replications Design (Quasi-Experimental)

Example for Diagram 4.6


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 adminis-
tered 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.)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>1 ((n = 65))</td>
<td>Preference for ideation, premature evaluations</td>
<td>Creative problem-solving program</td>
<td>Preference for ideation, premature evaluations</td>
<td>—</td>
<td>Preference for ideation, premature evaluations</td>
</tr>
<tr>
<td>NR</td>
<td>2 (wait-list; (n = 47))</td>
<td>Preference for ideation, premature evaluations</td>
<td>—</td>
<td>Preference for ideation, premature evaluations</td>
<td>Creative problem-solving program</td>
<td>Preference for ideation, premature evaluations</td>
</tr>
</tbody>
</table>

**Diagram 4.7** Crossover Design (2-Factor)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(O_1)</td>
<td>(X_A)</td>
<td>(O_2)</td>
<td>(X_B)</td>
<td>(O_3)</td>
</tr>
<tr>
<td>2</td>
<td>(O_1)</td>
<td>(X_B)</td>
<td>(O_2)</td>
<td>(X_A)</td>
<td>(O_3)</td>
</tr>
</tbody>
</table>
Example for Diagram 4.7


*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 ($M_{\text{age}} = 13.5$) 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

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 ($n = 25$)</td>
<td>BAQ, CY-PSPP, LTPAQ</td>
<td>Aerobic dance</td>
<td>BAQ, CY-PSPP, LTPAQ</td>
<td>Physical education</td>
<td>BAQ, CY-PSPP, LTPAQ</td>
</tr>
<tr>
<td>R</td>
<td>2 ($n = 25$)</td>
<td>BAQ, CY-PSPP, LTPAQ</td>
<td>Physical education</td>
<td>BAQ, CY-PSPP, LTPAQ</td>
<td>Aerobic dance</td>
<td>BAQ, CY-PSPP, LTPAQ</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretests</th>
<th>Treatment</th>
<th>Posttests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$O_1 \ldots O_8$</td>
<td>X</td>
<td>$O_7 \ldots O_{10}$</td>
</tr>
</tbody>
</table>

*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


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

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretests 1–8</th>
<th>Treatment</th>
<th>Posttests 7–10</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>1</td>
<td>VR-PICS</td>
<td>Verona Skills Training</td>
<td>VR-PICS</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretests</th>
<th>Treatment</th>
<th>Posttests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( O_1 \ldots O_6 )</td>
<td>X</td>
<td>( O_7 \ldots O_{11} )</td>
</tr>
<tr>
<td>2</td>
<td>( O_1 \ldots O_6 )</td>
<td></td>
<td>( O_7 \ldots O_{11} )</td>
</tr>
</tbody>
</table>

Time ▲
Chapter 4  Within-Subjects Approach  ♦  71

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


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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Years 1–6</th>
<th>Treatment</th>
<th>Years 7–11</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>1</td>
<td>SAT-9, CAT-5, DRP, NYS, PEP, PCT</td>
<td>America’s Choice</td>
<td>SAT-9, CAT-5, DRP, NYS, PEP, PCT</td>
</tr>
<tr>
<td>(Reading: $n = 56,693$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NR</td>
<td>2</td>
<td>SAT-9, CAT-5, DRP, NYS, PEP, PCT</td>
<td></td>
<td>SAT-9, CAT-5, DRP, NYS, PEP, PCT</td>
</tr>
<tr>
<td>(Math: $n = 55,932$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time ▼</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Diagram 4.10 6 x 6 Latin-Square Design

<table>
<thead>
<tr>
<th>Group</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A B C D E F</td>
</tr>
<tr>
<td>2</td>
<td>B C F A D E</td>
</tr>
<tr>
<td>3</td>
<td>C F B E A D</td>
</tr>
<tr>
<td>4</td>
<td>D E A B F C</td>
</tr>
<tr>
<td>5</td>
<td>E A D F C B</td>
</tr>
<tr>
<td>6</td>
<td>F D E C B A</td>
</tr>
</tbody>
</table>

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


Research Question: What are the effects of test 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)
### Chapter 4  Within-Subjects Approach

<table>
<thead>
<tr>
<th>Class (N = 135)</th>
<th>Stressful Instruction</th>
<th>Intelligence Tests</th>
<th>Reassuring Instruction</th>
<th>Intelligence Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RW-ES</td>
<td>Voc</td>
<td>Dice</td>
<td>RW-ES</td>
</tr>
<tr>
<td>2</td>
<td>RW-ES</td>
<td>Pic</td>
<td>Dice</td>
<td>RW-ES</td>
</tr>
<tr>
<td>3</td>
<td>RW-ES</td>
<td>Dice</td>
<td>Ana</td>
<td>RW-ES</td>
</tr>
<tr>
<td>4</td>
<td>RW-ES</td>
<td>Fig</td>
<td>Ana</td>
<td>RW-ES</td>
</tr>
<tr>
<td>5</td>
<td>RW-ES</td>
<td>Fig</td>
<td>Voc</td>
<td>RW-ES</td>
</tr>
<tr>
<td>6</td>
<td>RW-ES</td>
<td>Voc</td>
<td>Fig</td>
<td>RW-ES</td>
</tr>
</tbody>
</table>

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 $2 \times 2$ 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., $2 \times 2$ 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 $2 \times 2$ 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 $2 \times 3$ 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 $2 \times 3$ 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 [$2 \times 2$] and two posttest-only designs [$2 \times 2$ and $3 \times 2$]) and one example of a $2 \times 2$ 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:


**Diagram 5.1**

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( O_1 )</td>
<td>( X_{A1B1} )</td>
<td>( O_2 )</td>
</tr>
<tr>
<td>2</td>
<td>( O_1 )</td>
<td>( X_{A1B2} )</td>
<td>( O_2 )</td>
</tr>
<tr>
<td>3</td>
<td>( O_1 )</td>
<td>( X_{A2B1} )</td>
<td>( O_2 )</td>
</tr>
<tr>
<td>4</td>
<td>( O_1 )</td>
<td>( X_{A2B2} )</td>
<td>( O_2 )</td>
</tr>
</tbody>
</table>

Example for Diagram 5.1


*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 $2 \times 2$ 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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Factor Ability (A)</th>
<th>Factor Speech Type (B)</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1</td>
<td>Attitudes</td>
<td>Disabled ($A_1$)</td>
<td>Synthetic speech ($B_1$)</td>
<td>Attitudes, speech characteristics, effectiveness of message</td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>Attitudes</td>
<td>Disabled ($A_1$)</td>
<td>Normal speech ($B_2$)</td>
<td>Attitudes, speech characteristics, effectiveness of message</td>
</tr>
<tr>
<td>R</td>
<td>3</td>
<td>Attitudes</td>
<td>Nondisabled ($A_2$)</td>
<td>Synthetic speech ($B_1$)</td>
<td>Attitudes, speech characteristics, effectiveness of message</td>
</tr>
<tr>
<td>R</td>
<td>4</td>
<td>Attitudes</td>
<td>Nondisabled ($A_2$)</td>
<td>Normal speech ($B_2$)</td>
<td>Attitudes, speech characteristics, effectiveness of message</td>
</tr>
</tbody>
</table>

2x2 ($N = 189$)

<table>
<thead>
<tr>
<th>Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability (A)</td>
</tr>
<tr>
<td>Disabled ($A_1$)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech Type (B)</td>
</tr>
<tr>
<td>Computer-synthesized speech ($B_1$)</td>
</tr>
<tr>
<td>Natural speech ($B_2$)</td>
</tr>
</tbody>
</table>
Example for Diagram 5.2


**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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th><strong>Factor</strong></th>
<th><strong>Factor</strong></th>
<th><strong>Posttest</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Leader Type (A)</td>
<td>Speech Type (B)</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>1</td>
<td>Transformational (A₁)</td>
<td>Inspirational speech (B₁)</td>
<td>Multifactor Leadership Questionnaire</td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>Transformational (A₁)</td>
<td>Noninspirational speech (B₂)</td>
<td>Multifactor Leadership Questionnaire</td>
</tr>
<tr>
<td>R</td>
<td>3</td>
<td>Transactional (A₂)</td>
<td>Inspirational speech (B₁)</td>
<td>Multifactor Leadership Questionnaire</td>
</tr>
<tr>
<td>R</td>
<td>4</td>
<td>Transactional (A₂)</td>
<td>Noninspirational speech (B₂)</td>
<td>Multifactor Leadership Questionnaire</td>
</tr>
</tbody>
</table>

**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.
Diagram 5.3 2 x 2 Factorial Pretest and Posttest Design (Within-Subjects)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁</td>
<td>Xₐ₁β₁</td>
<td>O₂</td>
</tr>
<tr>
<td>1</td>
<td>O₁</td>
<td>Xₐ₁β₂</td>
<td>O₂</td>
</tr>
<tr>
<td>1</td>
<td>O₁</td>
<td>Xₐ₂β₁</td>
<td>O₂</td>
</tr>
<tr>
<td>1</td>
<td>O₁</td>
<td>Xₐ₂β₂</td>
<td>O₂</td>
</tr>
</tbody>
</table>

Time ➤

Example for Diagram 5.3


*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 x 2 factorial pretest and posttest design.
### Diagram 5.4 3 × 2 Factorial Posttest Design (Between-Subjects)

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X_{A_{1}B_{1}}</td>
<td>O_{1}</td>
</tr>
<tr>
<td>2</td>
<td>X_{A_{2}B_{1}}</td>
<td>O_{1}</td>
</tr>
<tr>
<td>3</td>
<td>X_{A_{3}B_{1}}</td>
<td>O_{1}</td>
</tr>
<tr>
<td>4</td>
<td>X_{A_{1}B_{2}}</td>
<td>O_{1}</td>
</tr>
<tr>
<td>5</td>
<td>X_{A_{2}B_{2}}</td>
<td>O_{1}</td>
</tr>
<tr>
<td>6</td>
<td>X_{A_{3}B_{2}}</td>
<td>O_{1}</td>
</tr>
</tbody>
</table>

*Time ➤*
**Recommended Parametric Analyses:** 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**


**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 $3 \times 2$ 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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Factor Highlighted (A)</th>
<th>Factor Text Difficulty (B)</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1</td>
<td>None ($A_1$)</td>
<td>Low ($B_1$)</td>
<td>Metacomprehension scale, comprehension test, confidence</td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>Appropriate ($A_2$)</td>
<td>Low ($B_1$)</td>
<td>Metacomprehension scale, comprehension test, confidence</td>
</tr>
<tr>
<td>R</td>
<td>3</td>
<td>Inappropriate ($A_3$)</td>
<td>Low ($B_1$)</td>
<td>Metacomprehension scale, comprehension test, confidence</td>
</tr>
<tr>
<td>R</td>
<td>4</td>
<td>None ($A_1$)</td>
<td>High ($B_2$)</td>
<td>Metacomprehension scale, comprehension test, confidence</td>
</tr>
<tr>
<td>R</td>
<td>5</td>
<td>Appropriate ($A_2$)</td>
<td>High ($B_2$)</td>
<td>Metacomprehension scale, comprehension test, confidence</td>
</tr>
<tr>
<td>R</td>
<td>6</td>
<td>Inappropriate ($A_3$)</td>
<td>High ($B_2$)</td>
<td>Metacomprehension scale, comprehension test, confidence</td>
</tr>
</tbody>
</table>

$2 \times 3$ ($N = 180$)

<table>
<thead>
<tr>
<th>Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highlighted Reading Passages (A)</td>
</tr>
<tr>
<td>Text Difficulty (B)</td>
</tr>
<tr>
<td>Low ($B_1$)</td>
</tr>
<tr>
<td>High ($B_2$)</td>
</tr>
</tbody>
</table>
### Example for Diagram 5.5


**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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Factor Instruction Type (A)</th>
<th>Factor Deictic Gesture (B)</th>
<th>Factor Facial Expression (C)</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1</td>
<td>Procedural (A₁)</td>
<td>Presence (B₁)</td>
<td>Presence (C₁)</td>
<td>API, attitudes, learning</td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>Procedural (A₁)</td>
<td>Absence (B₂)</td>
<td>Presence (C₁)</td>
<td>API, attitudes, learning</td>
</tr>
<tr>
<td>R</td>
<td>3</td>
<td>Attitudinal (A₂)</td>
<td>Presence (B₁)</td>
<td>Presence (C₁)</td>
<td>API, attitudes, learning</td>
</tr>
<tr>
<td>R</td>
<td>4</td>
<td>Attitudinal (A₂)</td>
<td>Absence (B₂)</td>
<td>Presence (C₁)</td>
<td>API, attitudes, learning</td>
</tr>
<tr>
<td>R</td>
<td>5</td>
<td>Procedural (A₁)</td>
<td>Presence (B₁)</td>
<td>Absence (C₂)</td>
<td>API, attitudes, learning</td>
</tr>
<tr>
<td>R</td>
<td>6</td>
<td>Procedural (A₁)</td>
<td>Absence (B₂)</td>
<td>Absence (C₂)</td>
<td>API, attitudes, learning</td>
</tr>
<tr>
<td>R</td>
<td>7</td>
<td>Attitudinal (A₂)</td>
<td>Presence (B₁)</td>
<td>Absence (C₂)</td>
<td>API, attitudes, learning</td>
</tr>
<tr>
<td>R</td>
<td>8</td>
<td>Attitudinal (A₂)</td>
<td>Absence (B₂)</td>
<td>Absence (C₂)</td>
<td>API, attitudes, learning</td>
</tr>
</tbody>
</table>

| Time ▶ |

<table>
<thead>
<tr>
<th>Independent Variable (IV)</th>
<th>N = 236</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction Type (A)</td>
<td></td>
</tr>
<tr>
<td>Procedural Module (A₁)</td>
<td>n = 29</td>
</tr>
<tr>
<td>Attitudinal Module (A₂)</td>
<td>n = 30</td>
</tr>
<tr>
<td>Facial Expression (C)</td>
<td></td>
</tr>
<tr>
<td>Presence (B₁)</td>
<td>n = 29</td>
</tr>
<tr>
<td>Absence (B₂)</td>
<td>n = 30</td>
</tr>
<tr>
<td>Presence (B₁)</td>
<td>n = 29</td>
</tr>
<tr>
<td>Absence (B₂)</td>
<td>n = 30</td>
</tr>
</tbody>
</table>
Example for Diagram 5.6


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.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Factor Legislative Action (A)</th>
<th>Factor Role of Evaluator (B)</th>
<th>Factor Accountability (C)</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1</td>
<td>Site-based (A₁)</td>
<td>Principal (B₃)</td>
<td>Low (C₁)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>Site-based (A₁)</td>
<td>Teacher (B₂)</td>
<td>Low (C₁)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>3</td>
<td>Site-based (A₁)</td>
<td>Parent (B₃)</td>
<td>Low (C₁)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>4</td>
<td>Nonsite-based (A₂)</td>
<td>Principal (B₃)</td>
<td>Low (C₁)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>5</td>
<td>Nonsite-based (A₂)</td>
<td>Teacher (B₂)</td>
<td>Low (C₁)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>6</td>
<td>Nonsite-based (A₂)</td>
<td>Parent (B₃)</td>
<td>Low (C₁)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>7</td>
<td>Site-based (A₁)</td>
<td>Principal (B₃)</td>
<td>High (C₂)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>8</td>
<td>Site-based (A₁)</td>
<td>Teacher (B₂)</td>
<td>High (C₂)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>9</td>
<td>Site-based (A₁)</td>
<td>Parent (B₃)</td>
<td>High (C₂)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>10</td>
<td>Nonsite-based (A₂)</td>
<td>Principal (B₃)</td>
<td>High (C₂)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>11</td>
<td>Nonsite-based (A₂)</td>
<td>Teacher (B₂)</td>
<td>High (C₂)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
<tr>
<td>R</td>
<td>12</td>
<td>Nonsite-based (A₂)</td>
<td>Parent (B₃)</td>
<td>High (C₂)</td>
<td>Perceived skill level, willingness to select candidate</td>
</tr>
</tbody>
</table>
### Chapter 5  Factorial Designs

#### 2 x 3 x 2

<table>
<thead>
<tr>
<th>IV</th>
<th>Role of Evaluator (B)</th>
<th>N = 624</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV</td>
<td>Legislative Action (A)</td>
<td>Principal (B₁)</td>
</tr>
<tr>
<td>Site-based (A₁)</td>
<td>n = 52</td>
<td>n = 52</td>
</tr>
<tr>
<td>Nonsite-based (A₂)</td>
<td>n = 52</td>
<td>n = 52</td>
</tr>
<tr>
<td>Site-based (A₃)</td>
<td>n = 52</td>
<td>n = 52</td>
</tr>
<tr>
<td>Nonsite-based (A₄)</td>
<td>n = 52</td>
<td>n = 52</td>
</tr>
</tbody>
</table>

#### Diagram 5.7  2 x 2 x 2 Mixed-Factorial Design (Mixed-Subjects)

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Xₐ₁B₁C₁C₂</td>
<td>O₁</td>
</tr>
<tr>
<td>2</td>
<td>Xₐ₁B₂C₁C₂</td>
<td>O₁</td>
</tr>
<tr>
<td>3</td>
<td>Xₐ₂B₁C₁C₂</td>
<td>O₁</td>
</tr>
<tr>
<td>4</td>
<td>Xₐ₂B₂C₁C₂</td>
<td>O₁</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Xₐ₁B₁C₂C₁</td>
<td>O₁</td>
</tr>
<tr>
<td>6</td>
<td>Xₐ₁B₂C₂C₁</td>
<td>O₁</td>
</tr>
<tr>
<td>7</td>
<td>Xₐ₂B₁C₂C₁</td>
<td>O₁</td>
</tr>
<tr>
<td>8</td>
<td>Xₐ₂B₂C₂C₁</td>
<td>O₁</td>
</tr>
</tbody>
</table>

**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 x 3 x 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**

### 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 \((n = 132 \text{ males}; \ n = 200 \text{ females})\) 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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Factor (BS)</th>
<th>Factor (BS)</th>
<th>Factor (WS)</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Infidelity Context (A)</td>
<td>Gender (B)</td>
<td>Infidelity Type (C)</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>1</td>
<td>Online (A₁)</td>
<td>Male (B₁)</td>
<td>Emotional (C₁) and sexual (C₂)</td>
<td>Test</td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>Online (A₁)</td>
<td>Female (B₂)</td>
<td>Emotional (C₁) and sexual (C₂)</td>
<td>Test</td>
</tr>
<tr>
<td>R</td>
<td>3</td>
<td>Conventional (A₂)</td>
<td>Male (B₁)</td>
<td>Emotional (C₁) and sexual (C₂)</td>
<td>Test</td>
</tr>
<tr>
<td>R</td>
<td>4</td>
<td>Conventional (A₂)</td>
<td>Female (B₂)</td>
<td>Emotional (C₁) and sexual (C₂)</td>
<td>Test</td>
</tr>
<tr>
<td>R</td>
<td>5</td>
<td>Online (A₁)</td>
<td>Male (B₁)</td>
<td>Sexual (C₂) and emotional (C₁)</td>
<td>Test</td>
</tr>
</tbody>
</table>
Chapter 5  Factorial Designs  ♦  91

### Table 5.1

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Factor (BS) Infidelity Context (A)</th>
<th>Factor (BS) Gender (B)</th>
<th>Factor (WS) Infidelity Type (C)</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>6</td>
<td>Online (A₁)</td>
<td>Female (B₂)</td>
<td>Sexual (C₁) and emotional (C₂)</td>
<td>Test</td>
</tr>
<tr>
<td>R</td>
<td>7</td>
<td>Conventional (A₂)</td>
<td>Male (B₁)</td>
<td>Sexual (C₁) and emotional (C₂)</td>
<td>Test</td>
</tr>
<tr>
<td>R</td>
<td>8</td>
<td>Conventional (A₂)</td>
<td>Female (B₂)</td>
<td>Sexual (C₁) and emotional (C₂)</td>
<td>Test</td>
</tr>
</tbody>
</table>

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

### Table 5.2

<table>
<thead>
<tr>
<th>2 x 2 x 2</th>
<th>Independent Variable (IV)</th>
<th>N = 332</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IV</strong></td>
<td>Infidelity Context (A)</td>
<td></td>
</tr>
<tr>
<td>Gender (B)</td>
<td>Online (A₁)</td>
<td>Conventional (A₂)</td>
</tr>
<tr>
<td>Male (B₁)</td>
<td>n = 33</td>
<td>n = 33</td>
</tr>
<tr>
<td>Female (B₂)</td>
<td>n = 50</td>
<td>n = 50</td>
</tr>
<tr>
<td>Male (B₁)</td>
<td>n = 33</td>
<td>n = 33</td>
</tr>
<tr>
<td>Female (B₂)</td>
<td>n = 50</td>
<td>n = 50</td>
</tr>
<tr>
<td>Male (B₁)</td>
<td>n = 33</td>
<td>n = 33</td>
</tr>
<tr>
<td>Female (B₂)</td>
<td>n = 50</td>
<td>n = 50</td>
</tr>
<tr>
<td>Male (B₁)</td>
<td>n = 33</td>
<td>n = 33</td>
</tr>
<tr>
<td>Female (B₂)</td>
<td>n = 50</td>
<td>n = 50</td>
</tr>
</tbody>
</table>

### 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:


**Diagram 6.1** Solomon Four-Group Design

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁</td>
<td>Xₐ</td>
<td>O₂</td>
</tr>
<tr>
<td>2</td>
<td>O₃</td>
<td>X₉</td>
<td>O₄</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>Xₐ</td>
<td>O₅</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>X₉</td>
<td>O₆</td>
</tr>
</tbody>
</table>

*Note*: It is highly recommended that random assignment be used when applying the Solomon N-group designs.
Example for Diagram 6.1


**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 x 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 x 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)
<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 (n = 64)</td>
<td>Survey</td>
<td>Organizational restructuring</td>
<td>Survey</td>
</tr>
<tr>
<td>R</td>
<td>2 (n = 62)</td>
<td>Survey</td>
<td>—</td>
<td>Survey</td>
</tr>
<tr>
<td>R</td>
<td>3 (n = 83)</td>
<td>—</td>
<td>Organizational restructuring</td>
<td>Survey</td>
</tr>
<tr>
<td>R</td>
<td>4 (n = 104)</td>
<td>—</td>
<td>—</td>
<td>Survey</td>
</tr>
</tbody>
</table>

**2 × 2 (n = 313)**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>n = 64 (O2)</td>
<td>n = 62 (O4)</td>
</tr>
<tr>
<td>No</td>
<td>n = 83 (O3)</td>
<td>n = 104 (O6)</td>
</tr>
</tbody>
</table>

**Diagram 6.2 Solomon Six-Group Design**

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁</td>
<td>Xₐ</td>
<td>O₂</td>
</tr>
<tr>
<td>2</td>
<td>O₃</td>
<td>X₉</td>
<td>O₄</td>
</tr>
<tr>
<td>3</td>
<td>—</td>
<td>Xₐ</td>
<td>O₅</td>
</tr>
<tr>
<td>4</td>
<td>—</td>
<td>X₉</td>
<td>O₆</td>
</tr>
<tr>
<td>5</td>
<td>O₇</td>
<td>—</td>
<td>O₈</td>
</tr>
<tr>
<td>6</td>
<td>—</td>
<td>—</td>
<td>O₉</td>
</tr>
</tbody>
</table>

**Note:** O₂, for example, represents the posttest observation for Group 1 from the design structure.

**Note:** The study had two treatment conditions (fake-good [Xₐ] and fake-bad [X₉]) 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


**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)
<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>1 (n = 50)</td>
<td>GMA, IPI, EIS</td>
<td>Fake-Good</td>
<td>GMA, IPI, EIS</td>
</tr>
<tr>
<td>R</td>
<td>2 (n = 50)</td>
<td>GMA, IPI, EIS</td>
<td>Fake-Bad</td>
<td>GMA, IPI, EIS</td>
</tr>
<tr>
<td>R</td>
<td>3 (n = 50)</td>
<td>—</td>
<td>Fake-Good</td>
<td>GMA, IPI, EIS</td>
</tr>
<tr>
<td>R</td>
<td>4 (n = 50)</td>
<td>—</td>
<td>Fake-Bad</td>
<td>GMA, IPI, EIS</td>
</tr>
<tr>
<td>R</td>
<td>5 (n = 50)</td>
<td>GMA, IPI, EIS</td>
<td>—</td>
<td>GMA, IPI, EIS</td>
</tr>
<tr>
<td>R</td>
<td>6 (n = 50)</td>
<td>—</td>
<td>—</td>
<td>GMA, IPI, EIS</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>2 x 2 (n = 200)</th>
<th>Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fake-Bad (X_b)</td>
</tr>
<tr>
<td>Independent Variable</td>
<td>Yes</td>
</tr>
<tr>
<td>Pretest</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>n = 50 (O_a)</td>
</tr>
<tr>
<td>No</td>
<td>n = 50 (O_a)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 x 2 (n = 200)</th>
<th>Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fake-Good (X_a)</td>
</tr>
<tr>
<td>Independent Variable</td>
<td>Yes</td>
</tr>
<tr>
<td>Pretest</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>n = 50 (O_2)</td>
</tr>
<tr>
<td>No</td>
<td>n = 50 (O_3)</td>
</tr>
</tbody>
</table>

Example for Diagram 6.3

**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)

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>1</td>
<td>Questionnaire</td>
<td>Lecture</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>NR</td>
<td>2</td>
<td>Questionnaire</td>
<td>Service Learning</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>NR</td>
<td>3</td>
<td>Questionnaire</td>
<td>Both</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>NR</td>
<td>4</td>
<td>—</td>
<td>Lecture</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>NR</td>
<td>5</td>
<td>—</td>
<td>Service Learning</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>NR</td>
<td>6</td>
<td>—</td>
<td>Both</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>NR</td>
<td>7</td>
<td>Questionnaire</td>
<td>—</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>NR</td>
<td>8</td>
<td>—</td>
<td>—</td>
<td>Questionnaire</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 x 2 (n = 210)</th>
<th>Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lecture (Xₐ)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Independent Variable</strong></td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Pretest</strong></td>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
<td>O₇</td>
</tr>
</tbody>
</table>
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.
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>
<thead>
<tr>
<th>Table 7.1</th>
<th>Design Notations for Single-Case Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Design Notation$</td>
<td>$Design Element$</td>
</tr>
<tr>
<td>$A$</td>
<td>Baseline</td>
</tr>
<tr>
<td>$B_A$</td>
<td>Treatment$_1$</td>
</tr>
<tr>
<td>$C$</td>
<td>Treatment$_n$</td>
</tr>
<tr>
<td>$O_n$</td>
<td>Observations</td>
</tr>
</tbody>
</table>

*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 (2019) 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:


### Diagram 7.1  A-B-A Design

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline A</th>
<th>Treatment B</th>
<th>Baseline A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(O_1 \ldots O_4)</td>
<td>(O_1 \ldots O_4)</td>
<td>(O_1 \ldots O_5)</td>
</tr>
</tbody>
</table>

**Example for Diagram 7.1**


**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)

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline (A) Four Observations</th>
<th>Treatment (B) Four Observations</th>
<th>Baseline (A) Five Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disruptive behavior, correct response, on-task behavior</td>
<td>Disruptive behavior, correct response, on-task behavior</td>
<td>Disruptive behavior, correct response, on-task behavior</td>
</tr>
</tbody>
</table>

Diagram 7.2 A-B-A-B Design

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline A</th>
<th>Treatment B</th>
<th>Baseline A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁ ... O₁₀</td>
<td>O₁ ... O₁₀</td>
<td>O₁ ... O₁₃</td>
<td>O₁ ... O₁₀</td>
</tr>
<tr>
<td>2</td>
<td>O₁ ... O₁₀</td>
<td>O₁ ... O₁₀</td>
<td>O₁ ... O₁₃</td>
<td>O₁ ... O₁₀</td>
</tr>
<tr>
<td>3</td>
<td>O₁ ... O₁₀</td>
<td>O₁ ... O₁₀</td>
<td>O₁ ... O₁₃</td>
<td>O₁ ... O₁₀</td>
</tr>
</tbody>
</table>

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


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)

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline (A) 10 Observations</th>
<th>Treatment (B) ~10 Observations</th>
<th>Baseline (A) 13 Observations</th>
<th>Treatment (B) ~10 Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OMNI Scale, exercise diary</td>
<td>OMNI Scale, exercise diary</td>
<td>OMNI Scale, exercise diary</td>
<td>OMNI Scale, exercise diary</td>
</tr>
<tr>
<td>2</td>
<td>OMNI Scale, exercise diary</td>
<td>OMNI Scale, exercise diary</td>
<td>OMNI Scale, exercise diary</td>
<td>OMNI Scale, exercise diary</td>
</tr>
<tr>
<td>3</td>
<td>OMNI Scale, exercise diary</td>
<td>OMNI Scale, exercise diary</td>
<td>OMNI Scale, exercise diary</td>
<td>OMNI Scale, exercise diary</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline</th>
<th>Treatment A</th>
<th>Baseline</th>
<th>Treatment A</th>
<th>Treatment B</th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
</tr>
<tr>
<td>2</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
</tr>
<tr>
<td>3</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
<td>Oₙ</td>
</tr>
</tbody>
</table>

Note: Each Oₙ represents multiple measures (observations), sometimes referred to as sessions.
Example for Diagram 7.3


**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)

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline (A) Five Observations</th>
<th>Treatment$_A$ (B) Five Observations</th>
<th>Baseline (A) Six Observations</th>
<th>Treatment$_A$ (B) Six Observations</th>
<th>Treatment$_B$ (C) Five Observations</th>
<th>Treatment$_A$ (B) Five Observations</th>
<th>Treatment$_B$ (C) Six Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
</tr>
<tr>
<td>2</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
</tr>
<tr>
<td>3</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
<td>Behavior</td>
</tr>
</tbody>
</table>
Diagram 7.4 Changing Criterion Design (A-B)

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline (A)</th>
<th>Treatment (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁ … O₅</td>
<td>O₁ … O₂₀</td>
</tr>
<tr>
<td>2</td>
<td>O₁ … O₅</td>
<td>O₁ … O₂₀</td>
</tr>
<tr>
<td>3</td>
<td>O₁ … O₅</td>
<td>O₁ … O₂₀</td>
</tr>
</tbody>
</table>

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

Example for Diagram 7.4


**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 ($M_{age} = 10.5$; $N = 3$) 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, interresponse time, and celeration)
Chapter 7  Single-Case Approach ♦ 111

### Example for Diagram 7.5


**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 acceleration)

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline (A)</th>
<th>Treatment (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~5 Observations</td>
<td>~24 Observations</td>
</tr>
<tr>
<td>1</td>
<td>CVI-IV, engagement, behavior</td>
<td>CVI-IV, engagement, behavior</td>
</tr>
<tr>
<td>2</td>
<td>Baseline (A)</td>
<td>Treatment (B)</td>
</tr>
<tr>
<td></td>
<td>~9 Observations</td>
<td>~15 Observations</td>
</tr>
<tr>
<td>2</td>
<td>CVI-IV, engagement, behavior</td>
<td>CVI-IV, engagement, behavior</td>
</tr>
<tr>
<td>3</td>
<td>Baseline (A)</td>
<td>Treatment (B)</td>
</tr>
<tr>
<td></td>
<td>~13 Observation</td>
<td>~16 Observations</td>
</tr>
<tr>
<td>3</td>
<td>CVI-IV, engagement, behavior</td>
<td>CVI-IV, engagement, behavior</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline A&lt;sub&gt;i&lt;/sub&gt;</th>
<th>Treatment&lt;sub&gt;A&lt;/sub&gt; B</th>
<th>Treatment&lt;sub&gt;g&lt;/sub&gt; C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁ ... O&lt;sub&gt;n&lt;/sub&gt;</td>
<td>O₁ ... O&lt;sub&gt;n&lt;/sub&gt;</td>
<td>O₁ ... O&lt;sub&gt;n&lt;/sub&gt;</td>
</tr>
<tr>
<td>2</td>
<td>O₁ ... O&lt;sub&gt;n&lt;/sub&gt;</td>
<td>O₁ ... O&lt;sub&gt;n&lt;/sub&gt;</td>
<td>O₁ ... O&lt;sub&gt;n&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

**Note:** Four cases were used in this study.
Example for Diagram 7.6


**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) treatmentA, (c) treatmentB, 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

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline (A)</th>
<th>Treatment (B)</th>
<th>Treatment (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~3 Observations</td>
<td>~11 Observations</td>
<td>~10 Observations</td>
</tr>
<tr>
<td>1</td>
<td>DPICS</td>
<td>DPICS</td>
<td>DPICS</td>
</tr>
<tr>
<td>2</td>
<td>Baseline (A)</td>
<td>Treatment (B)</td>
<td>Treatment (C)</td>
</tr>
<tr>
<td></td>
<td>~6 Observations</td>
<td>~9 Observations</td>
<td>~10 Observations</td>
</tr>
<tr>
<td>3</td>
<td>Baseline (A)</td>
<td>Treatment (B)</td>
<td>Treatment (C)</td>
</tr>
<tr>
<td></td>
<td>~9 Observations</td>
<td>~13 Observations</td>
<td>~6 Observations</td>
</tr>
<tr>
<td>4</td>
<td>Baseline (A)</td>
<td>Treatment (B)</td>
<td>Treatment (C)</td>
</tr>
<tr>
<td></td>
<td>~11 Observations</td>
<td>~13 Observations</td>
<td>~7 Observations</td>
</tr>
</tbody>
</table>

*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, interre­sponse 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

<table>
<thead>
<tr>
<th>Method</th>
<th>Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>Nonexperimental</td>
</tr>
<tr>
<td>Approach</td>
<td>Observational</td>
</tr>
<tr>
<td>Design</td>
<td>Posttest (one-group) Ex post facto</td>
</tr>
</tbody>
</table>

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).
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

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
<th>Assignment</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>O&lt;sub&gt;1&lt;/sub&gt;</td>
<td>M&lt;sub&gt;A&lt;/sub&gt;</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

*Note:* M<sub>A</sub> 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)

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>O&lt;sub&gt;1&lt;/sub&gt;</td>
</tr>
</tbody>
</table>

*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


*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

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>1 (N = 151)</td>
<td>Tobacco hazard education</td>
<td>Tobacco use, intent to quit</td>
</tr>
</tbody>
</table>

![Diagram 8.3](Ex Post Facto Design)

*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

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)

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
<th>Assignment</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N = 67$</td>
<td>DVRS program</td>
<td>PANAS, SWLS, CES-D</td>
<td>Unsuccessful</td>
<td>1 ($n = 31$)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PANAS, SWLS, CES-D</td>
<td>Successful</td>
<td>2 ($n = 36$)</td>
</tr>
</tbody>
</table>

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
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.
PART II QUANTITATIVE METHODS FOR NONEXPERIMENTAL RESEARCH

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 multilevel linear modeling, contextual modeling, mixed models, discriminant function analysis, and principal and exploratory component analysis.

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


**Example for Diagram 9.1**


*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
### Variable | Observation
--- | ---
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 Design

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observation</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor</td>
<td>O₁</td>
<td>—</td>
</tr>
<tr>
<td>Criterion</td>
<td>—</td>
<td>O₁</td>
</tr>
</tbody>
</table>

*Time ➤*

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

**Example for Diagram 9.2**


*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 | Variable | Observation | Variable | Observation | Variable | Observation | Variable | Observation
---|---|---|---|---|---|---|---|---|---|---
Mental factors | KAI, PSI, GSTB, CAS | Programming achievement | | | | | | | |

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

**Diagram 9.3** Predictive Design

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observation</th>
<th>Observation</th>
<th>Observation</th>
<th>Observation</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor</td>
<td>O₁</td>
<td>O₂</td>
<td>O₃</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>O₁</td>
<td>O₂</td>
<td>O₃</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example for Diagram 9.3**


*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
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.
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).
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 nonresponse 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).
We refer the reader to the following book for further details regarding survey approaches:


**Example for Diagram 10.1**


*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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-back pain</td>
<td>Low-back pain survey</td>
</tr>
</tbody>
</table>

**Diagram 10.2** Longitudinal Design

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observation</th>
<th>Observation</th>
<th>Observation</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$O_1$</td>
<td>$O_2$</td>
<td>$O_3$</td>
<td>$O_4$</td>
</tr>
<tr>
<td>2</td>
<td>$O_1$</td>
<td>$O_2$</td>
<td>$O_3$</td>
<td>$O_4$</td>
</tr>
<tr>
<td>3</td>
<td>$O_1$</td>
<td>$O_2$</td>
<td>$O_3$</td>
<td>$O_4$</td>
</tr>
<tr>
<td>4</td>
<td>$O_1$</td>
<td>$O_2$</td>
<td>$O_3$</td>
<td>$O_4$</td>
</tr>
<tr>
<td>5</td>
<td>$O_1$</td>
<td>$O_2$</td>
<td>$O_3$</td>
<td>$O_4$</td>
</tr>
</tbody>
</table>

*Note:* Any number of variables and observations can be associated with this design.
Chapter 10  Survey Approach ♦ 137

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

*Recommended Statistical Analysis:* Descriptive statistics

**Example for Diagram 10.2**


*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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observation</th>
<th>Observation</th>
<th>Observation</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Full Sample</strong></td>
<td><strong>Panel</strong></td>
<td><strong>Random Sample</strong></td>
<td><strong>Panel</strong></td>
</tr>
<tr>
<td></td>
<td>(n = 288)</td>
<td>(n = 92)</td>
<td>(n = 481)</td>
<td>(n = 92)</td>
</tr>
<tr>
<td>Internet use</td>
<td>Internet use survey</td>
<td>Internet use survey</td>
<td>Internet use survey</td>
<td>Internet use survey</td>
</tr>
<tr>
<td>Social network use</td>
<td>Social network survey</td>
<td>Social network survey</td>
<td>Social network survey</td>
<td>Social network survey</td>
</tr>
<tr>
<td>Well-being</td>
<td>Well-being survey</td>
<td>Well-being survey</td>
<td>Well-being survey</td>
<td>Well-being survey</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>Self-esteem survey</td>
<td>Self-esteem survey</td>
<td>Self-esteem survey</td>
<td>Self-esteem survey</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Satisfaction survey</td>
<td>Satisfaction survey</td>
<td>Satisfaction survey</td>
<td>Satisfaction survey</td>
</tr>
<tr>
<td></td>
<td><strong>Time ▶</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th>Method</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>Nonexperimental</td>
</tr>
<tr>
<td>Perspective</td>
<td>Grounded Theory</td>
</tr>
<tr>
<td>Design</td>
<td>Systematic</td>
</tr>
<tr>
<td></td>
<td>Descriptive</td>
</tr>
</tbody>
</table>

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:


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:


**Figure 11.1** Systematic Design

![Systematic Design Diagram]

**Example for Figure 11.1**


*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

Fit
Do concepts fit with the incidents they are representing?

- -

Relevance
The real concerns of the participants are represented.

- -

Workability
The variations in participants' behaviors are explained.

- -

Modifiability
Compare new data to existing data to improve theory.

Grounded Theory

Source: Glaser and Strauss (1967).
**Design:** Qualitative method using a grounded theory perspective with a systematic design

### Example for Figure 11.2


**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.
Example for Figure 11.3


*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.
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:


Example for Figure 12.1


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
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

Collaborate with participants to prevent further marginalization (give back to the participants studied)

Empower participants through “speaking” on their behalf (nonneutral stance)

Goal
To change societal views or challenge the status quo

Embrace biases in research (researcher’s role is reflexive and self-aware)

Connect the meaning of a situation to the broader context of social issues

---

**Example for Figure 12.2**


**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 12.3**


![Figure 12.3 Case Study Design](image)

---

According 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

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).

**DESCRIPTION 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 frame­
work 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


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 lan­
guage 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 lan­
guage and literacy from the preschool years through the end of adoles­
cence. The participants were asked to recall personally significant events,
situations, and people that made a difference to their learning, as well as
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

---

**Example for Figure 13.2**


*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

- **Stage 1**: Identify a system or situation to critique.
- **Stage 2**: Use purposeful sampling.
- **Stage 3**: Collect stories through collaborating with participants (to prevent further marginalization).
- **Stage 4**: Restory or retell. Embrace biases in research (researcher’s role is reflexive and self-aware).
- **Stage 5**: Collaborate with the participants.
- **Stage 6**: Write a story about the participants’ experiences. Empower participants through “speaking” on their behalf (nonneutral stance).
- **Stage 7**: Validate the accuracy of the narrative account. Connect the meaning of the situation to the broader context of social issues.
Design: Qualitative method using a narrative perspective with an explanatory design

Example for Figure 13.3


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
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.
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:


Vagle, M. D. (2014). *Crafting phenomenological research*. Walnut Creek, CA: Left Coast Press.


**Example for Figure 14.1**


**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.
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
Example for Figure 14.2


**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 *epoche*, 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 (*epoche* 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

**Example for Figure 14.3**

**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.

---

**Example for Figure 14.4**


---

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 disabili-

**Research Aim:** To understand the impact of placement and special educa-
tion services, especially in relation to receiving services in inclusive general
education classes

**Procedures:** The study used a phenomenological lens to explore the experi-
cence 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 sup-
ports 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 partici-
pant'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 expe-
riences. 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 mak-
ing 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

<table>
<thead>
<tr>
<th>Method</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research</td>
<td>Experimental</td>
</tr>
<tr>
<td>Approach</td>
<td>Convergence-Parallel</td>
</tr>
<tr>
<td>Design</td>
<td>Parallel-Databases</td>
</tr>
<tr>
<td></td>
<td>Embedded-Correlational</td>
</tr>
<tr>
<td></td>
<td>Participant-Selection</td>
</tr>
</tbody>
</table>

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 legitimatize 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, 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:

<table>
<thead>
<tr>
<th>Design Notation</th>
<th>Design Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUAN</td>
<td>Quantitatively driven study</td>
</tr>
<tr>
<td>QUAL</td>
<td>Qualitatively driven study</td>
</tr>
<tr>
<td>quan</td>
<td>Quantitative data is secondary to qualitative data</td>
</tr>
<tr>
<td>qual</td>
<td>Qualitative data is secondary to quantitative data</td>
</tr>
<tr>
<td>+</td>
<td>Indicates that quantitative and qualitative data are collected concurrently</td>
</tr>
<tr>
<td>→</td>
<td>Indicates that quantitative and qualitative data are collected sequentially</td>
</tr>
<tr>
<td>()</td>
<td>Parentheses indicate that one method is embedded within an emphasized method such as QUAL(quant)</td>
</tr>
<tr>
<td>⇛</td>
<td>Indicates that the methods are implemented recursively</td>
</tr>
<tr>
<td>[ ]</td>
<td>Brackets indicate a mixed method study that is within a series of studies</td>
</tr>
<tr>
<td>=</td>
<td>Indicates the transition to mixing methods</td>
</tr>
</tbody>
</table>

*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:


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.
PART IV  MIXED METHODS

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.
Example for Figure 15.1


**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 design.\(^1\)

---

**Figure 15.2** Data-Transformation Design

![Data-Transformation Design Diagram](image)

*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**


---

\(^1\)The authors refer to the design as a concurrent triangulation strategy, which follows the parallel-databases design.
Research Questions

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.

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


**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.
Example for Figure 15.4


**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.
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.
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 16.1


**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...
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](image)

**Figure 16.2** Embedded-Correlational Design

**QUAN**  
(IV or predictor variable)  

**QUAN**  
(DV or criterion variable)  

Interpret  
**QUAN**  
(qual)  

Collect  
qual  

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**


**Research Questions**

*Quantitative and Qualitative:* Are individuals with schizophrenia motivated to smoke for relief of psychiatric symptoms and to relieve antipsychotic 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](image_url)  
**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


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.
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

| QUAN data collection, analysis, and results | Identify results for follow-up | qual data collection | qual data analysis and results | Interpret QUAN → qual |

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


*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

| quan data collection, analysis, and results | QUAL participant selection | QUAL data collection | QUAL data analysis and results | Interpret quan → QUAL |

*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.**


**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.
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 subse-
quent quantitative analysis (Phase 2). It is also a useful approach for revis-
ing 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 pro-
vides 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 dur-
ing 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

```
qual data collection, analysis, and results → Develop instruments → QUAN data collection → QUAN data analysis and results → Interpret qual → QUAN
```

*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**


**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**


*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

![Treatment-Development Design](image)

*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**


**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

<table>
<thead>
<tr>
<th>Case</th>
<th>Method</th>
<th>Baseline A</th>
<th>Treatment B</th>
<th>Baseline A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>QUAN</td>
<td>O_n</td>
<td>O_n</td>
<td>O_n</td>
</tr>
<tr>
<td></td>
<td>qual</td>
<td>O_qual</td>
<td>O_qual</td>
<td>O_qual</td>
</tr>
</tbody>
</table>

*Note: Multiple forms and types of qual data can be collected during each phase.*
Diagram 19.2  A-B-A Sequential Design

<table>
<thead>
<tr>
<th>Case</th>
<th>Baseline A</th>
<th>Treatment B</th>
<th>Baseline A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$O_n$</td>
<td>$O_{\text{qual}}$</td>
<td>$O_n$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$O_{\text{qual}}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$O_n$</td>
</tr>
</tbody>
</table>

Time $\rightarrow$

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

<table>
<thead>
<tr>
<th>Case</th>
<th>Pretest</th>
<th>Baseline A</th>
<th>Treatment B</th>
<th>Baseline A</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$O_{\text{qual}}$</td>
<td>$O_n$</td>
<td>$O_n$</td>
<td>$O_n$</td>
<td>$O_{\text{qual}}$</td>
</tr>
</tbody>
</table>

Time $\rightarrow$

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.
he 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
We refer the reader to the following books to learn more about the action research approach:


**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 ($O_i$) 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

<table>
<thead>
<tr>
<th>Design Notation</th>
<th>Design Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_n$</td>
<td>The dependent variable (includes qual and quan data collected from artifact, observational and inquiry data)</td>
</tr>
<tr>
<td>$X$</td>
<td>The independent variable—also known as the treatment, factor, or program</td>
</tr>
<tr>
<td>Analyze</td>
<td>Analyze qualitative and quantitative data</td>
</tr>
<tr>
<td>Feedback</td>
<td>Based on the analyses, researcher makes necessary adjustments, feeds findings back into program</td>
</tr>
<tr>
<td>Evaluate</td>
<td>After feedback adjustments, program is evaluated, tested again</td>
</tr>
</tbody>
</table>

### Diagram 20.1 Posttest With a Historical Control for the Action Research Approach

<table>
<thead>
<tr>
<th>Group</th>
<th>Test</th>
<th>Program</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$O_1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$X$</td>
<td>$O_1$</td>
<td></td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Group</th>
<th>Test</th>
<th>Program</th>
<th>Posttest</th>
<th>Analyze</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Evaluate

Diagram 20.3 Posttest-Only Design for the Action Research Approach

<table>
<thead>
<tr>
<th>Group</th>
<th>Program</th>
<th>Posttest</th>
<th>Analyze</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>O₁</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Evaluate

Diagram 20.4 Ex Post Facto Design for the Action Research Approach

<table>
<thead>
<tr>
<th>Program already occurred</th>
<th>Group</th>
<th>Program</th>
<th>Posttest</th>
<th>Analyze</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td></td>
<td>O₁</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Evaluate

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.
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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Proxy Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>O_{A1}</td>
<td>X</td>
<td>O_{B1}</td>
</tr>
<tr>
<td>2</td>
<td>NR</td>
<td>O_{A1}</td>
<td>—</td>
<td>O_{B1}</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>O_{1}</td>
<td>O_{2}</td>
<td>X</td>
<td>O_{3}</td>
</tr>
<tr>
<td>2</td>
<td>NR</td>
<td>O_{1}</td>
<td>O_{2}</td>
<td>—</td>
<td>O_{3}</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>$O_1$</td>
<td>$O_2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>NR</td>
<td>-</td>
<td>-</td>
<td>$O_1$</td>
<td>$X$</td>
<td>$O_2$</td>
</tr>
</tbody>
</table>

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 $O_1$ to $O_2$ from the historical control to $O_1$ to $O_2$ 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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Test</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>$O_1$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>NR</td>
<td>-</td>
<td>$X$</td>
<td>$O_1$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Test</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R</td>
<td>-</td>
<td>$O_1$</td>
<td>$X$</td>
<td>$O_2$</td>
</tr>
<tr>
<td>2</td>
<td>R</td>
<td>-</td>
<td>$O_1$</td>
<td>-</td>
<td>$O_2$</td>
</tr>
<tr>
<td>3</td>
<td>NR</td>
<td>$O_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Unit</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>$O_1$</td>
<td>X</td>
<td>$O_2$</td>
</tr>
<tr>
<td>2*</td>
<td>NR</td>
<td>$O_1$</td>
<td></td>
<td>$O_2$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>$O_1$</td>
<td>$O_2$</td>
<td>X</td>
<td>$O_3$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>$O_1$</td>
<td>X</td>
<td>$O_2$</td>
<td>$O_3$</td>
<td></td>
<td>$O_4$</td>
</tr>
</tbody>
</table>

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 $O_1$ 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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NR</td>
<td>O₁</td>
<td>X</td>
<td>O₂</td>
<td>—</td>
<td>O₃</td>
<td>X</td>
<td>O₄</td>
</tr>
</tbody>
</table>

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 O₄. 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

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Midtest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R</td>
<td>O₁</td>
<td>Xᴬ</td>
<td>O₂</td>
<td>Xᴮ</td>
<td>O₃</td>
<td>Xᶜ</td>
<td>O₄</td>
</tr>
<tr>
<td>2</td>
<td>R</td>
<td>O₁</td>
<td>Xᴮ</td>
<td>O₂</td>
<td>Xᴬ</td>
<td>O₃</td>
<td>Xᶜ</td>
<td>O₄</td>
</tr>
<tr>
<td>3</td>
<td>R</td>
<td>O₁</td>
<td>Xᶜ</td>
<td>O₂</td>
<td>Xᴬ</td>
<td>O₃</td>
<td>Xᴮ</td>
<td>O₄</td>
</tr>
</tbody>
</table>

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 x 3 Graeco-Latin Square Design**

<table>
<thead>
<tr>
<th>Blocking Factor (Levels 1–3)</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A_a</td>
<td>B_y</td>
<td>C_p</td>
</tr>
<tr>
<td>2</td>
<td>B_b</td>
<td>C_a</td>
<td>A_y</td>
</tr>
<tr>
<td>3</td>
<td>C_y</td>
<td>A_b</td>
<td>B_a</td>
</tr>
</tbody>
</table>

*Note:* This is an example of a $3 \times 3$ 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 $\alpha, \gamma,$ and $\beta$. 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 x 3 Repeated Latin-Square Design**

<table>
<thead>
<tr>
<th>Group</th>
<th>Blocking Factor 1</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Level 1</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>Level 2</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>Level 3</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

*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—\( 3^1 \times 3^2 = 9 \))

<table>
<thead>
<tr>
<th>Group</th>
<th>Anxiety</th>
<th>Putt Difficulty</th>
<th>Putt Difficulty</th>
<th>Putt Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n = 9)</td>
<td>High</td>
<td>Easy</td>
<td>Medium</td>
<td>Hard</td>
</tr>
<tr>
<td>2 (n = 9)</td>
<td>Medium</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>3 (n = 9)</td>
<td>Low</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Day 1

<table>
<thead>
<tr>
<th>Group</th>
<th>Anxiety</th>
<th>Putt Difficulty</th>
<th>Putt Difficulty</th>
<th>Putt Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n = 9)</td>
<td>High</td>
<td>Easy</td>
<td>Medium</td>
<td>Hard</td>
</tr>
<tr>
<td>2 (n = 9)</td>
<td>Medium</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>3 (n = 9)</td>
<td>Low</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Day 2

<table>
<thead>
<tr>
<th>Group</th>
<th>Anxiety</th>
<th>Putt Difficulty</th>
<th>Putt Difficulty</th>
<th>Putt Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n = 9)</td>
<td>High</td>
<td>Easy</td>
<td>Medium</td>
<td>Hard</td>
</tr>
<tr>
<td>2 (n = 9)</td>
<td>Medium</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>3 (n = 9)</td>
<td>Low</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic</td>
<td>The examination of a unique case</td>
</tr>
<tr>
<td>Instrumental</td>
<td>The examination of a case to provide insight into an issue or specific theme</td>
</tr>
<tr>
<td>Multiple Instrumental (also known as Collective Case Study)</td>
<td>The examination approach, which is the same as instrumental, but with multiple cases</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Case–Holistic</td>
<td></td>
</tr>
<tr>
<td>Critical case</td>
<td>Examine a well-formulated theory.</td>
</tr>
<tr>
<td>Unique case</td>
<td>Examine an extreme case.</td>
</tr>
<tr>
<td>Representative case</td>
<td>Examine a typical or average case.</td>
</tr>
<tr>
<td>Revelatory case</td>
<td>Examine a phenomenon that was previously inaccessible.</td>
</tr>
<tr>
<td>Longitudinal case</td>
<td>Examine the same cover over a period of time.</td>
</tr>
<tr>
<td>Single Case–Embedded</td>
<td></td>
</tr>
<tr>
<td>Critical case</td>
<td>Examine a well-formulated theory.</td>
</tr>
<tr>
<td>Unique case</td>
<td>Examine an extreme case.</td>
</tr>
<tr>
<td>Representative case</td>
<td>Examine a typical or average case.</td>
</tr>
<tr>
<td>Revelatory case</td>
<td>Examine a phenomenon that was previously inaccessible.</td>
</tr>
<tr>
<td>Longitudinal case</td>
<td>Examine the same cover over a period of time.</td>
</tr>
<tr>
<td>Multiple Case–Holistic</td>
<td></td>
</tr>
<tr>
<td>Literal replication</td>
<td>Select and examine each case so that all cases are presumed to predict similar results.</td>
</tr>
<tr>
<td>Theoretical replication</td>
<td>Select and examine each case so that all cases are presumed to predict contrasting results, but for anticipatable reasons.</td>
</tr>
<tr>
<td>Multiple Case–Embedded</td>
<td></td>
</tr>
<tr>
<td>Literal replication</td>
<td>Select and examine each case so that all cases are presumed to predict similar results.</td>
</tr>
<tr>
<td>Theoretical replication</td>
<td>Select and examine each case so that all cases are presumed to predict contrasting results, but for anticipatable reasons.</td>
</tr>
</tbody>
</table>
## Appendix C

### Five Types of Mixed Methods Designs

<table>
<thead>
<tr>
<th>Design</th>
<th>Procedures</th>
</tr>
</thead>
</table>
| Parallel mixed  | • 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| • 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| • 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| • 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| • 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

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Statistical Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Interval IV</td>
<td>Interval</td>
<td>1) Correlation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) Simple linear regression</td>
</tr>
<tr>
<td></td>
<td>Ordinal</td>
<td>2) Nonparametric correlation</td>
</tr>
<tr>
<td></td>
<td>Nominal</td>
<td>3) Simple logistic regression</td>
</tr>
</tbody>
</table>

### Statistical Tests for One or More Interval IV and/or One or More Categorical IV

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Statistical Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>One or more Interval IV and/or one or more Categorical IV</td>
<td>Interval</td>
<td>1) Multiple regression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2) Analysis of covariance</td>
</tr>
<tr>
<td></td>
<td>Categorical</td>
<td>3) Multiple logistic regression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4) Discriminant analysis</td>
</tr>
</tbody>
</table>

### INFERENTIAL

### Statistical Tests for Zero Independent Variables

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Statistical Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero IVs (One sample/group)</td>
<td>Interval and existing value</td>
<td>1) One-sample ( t ) test</td>
</tr>
<tr>
<td></td>
<td>Ordinal and existing value</td>
<td>2) One-sample median</td>
</tr>
<tr>
<td></td>
<td>Nominal (two categories) and existing value</td>
<td>3) Binomial test</td>
</tr>
<tr>
<td></td>
<td>Nominal and existing value</td>
<td>4) Chi-square goodness of fit</td>
</tr>
</tbody>
</table>
### Statistical Tests for One Independent Variable

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Statistical Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>One IV with two levels (Independent groups)</td>
<td>Interval</td>
<td>1) Independent sample t test</td>
</tr>
<tr>
<td></td>
<td>Ordinal</td>
<td>2) Wilcoxon-Mann Whitney</td>
</tr>
<tr>
<td></td>
<td>Nominal (2 categories)</td>
<td>3.1) Chi-square 3.2) Fisher’s exact</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Statistical Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>One IV with two or more levels (Independent groups)</td>
<td>Interval</td>
<td>1) One-way ANOVA</td>
</tr>
<tr>
<td></td>
<td>Ordinal</td>
<td>2) Kruskal Wallis</td>
</tr>
<tr>
<td></td>
<td>Categorical</td>
<td>3) Chi-square</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Statistical Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>One IV with two levels (Dependent/paired groups)</td>
<td>Interval</td>
<td>1) Paired samples t test</td>
</tr>
<tr>
<td></td>
<td>Ordinal</td>
<td>2) Wilcoxon signed ranks</td>
</tr>
<tr>
<td></td>
<td>Categorical</td>
<td>3) McNemar</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Statistical Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>One IV with two or more levels (Dependent/paired groups)</td>
<td>Interval</td>
<td>1) One-way repeated measures ANOVA</td>
</tr>
<tr>
<td></td>
<td>Ordinal</td>
<td>2) Friedman test</td>
</tr>
<tr>
<td></td>
<td>Categorical</td>
<td>3) Logistic regression</td>
</tr>
</tbody>
</table>

### Statistical Tests for Two or More Independent Variables

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Statistical Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 or more IVs (Independent groups)</td>
<td>Interval</td>
<td>1) 2- or 3-way ANOVA</td>
</tr>
<tr>
<td></td>
<td>Categorical</td>
<td>2) Logistic regression</td>
</tr>
</tbody>
</table>
Statistical Tests for One Independent Variable
With Two or More Levels

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Dependent Variable</th>
<th>Statistical Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>One IV with two or more levels (independent group)</td>
<td>Interval</td>
<td>1) One-way MANOVA</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Sadness</th>
<th>Anger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger Adults (18–28)</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>T1</td>
<td>1.66Aa</td>
<td>1.37Aa</td>
</tr>
<tr>
<td>SD</td>
<td>0.90</td>
<td>0.84</td>
</tr>
<tr>
<td>T2</td>
<td>3.41Ba</td>
<td>3.43Ba</td>
</tr>
<tr>
<td>SD</td>
<td>1.06</td>
<td>1.15</td>
</tr>
<tr>
<td>Older Adults (62–87)</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>T1</td>
<td>1.38Ab</td>
<td>1.19Aa</td>
</tr>
<tr>
<td>SD</td>
<td>0.68</td>
<td>0.59</td>
</tr>
<tr>
<td>T2</td>
<td>3.10Bb</td>
<td>3.30Ba</td>
</tr>
<tr>
<td>SD</td>
<td>1.19</td>
<td>1.30</td>
</tr>
</tbody>
</table>

Source: Hofer et al., 2015, p. 51.

The following from Reynolds, McCambridge, Bissett, and Consedine (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' $\Lambda = .90$, $F(2, 77) = 4.40, p = .016$, $\eta^2_p = .10$; the experimental group reported more disgust than the
control group at both the beginning, $F(1, 78) = 5.65, p = .020$, $\eta^2 = .07$, and end of the session, $F(1, 78) = 7.75, p = .007, \eta^2 = .09$. 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' $\Lambda = 1.00, F(1, 78) = .11, p = .739, \eta^2 = .001$. Importantly, in terms of eliminating other possible affective bases for avoidance, the induction was specific to disgust, as embarrassment, $F(2, 77) = .12, p = .884, \eta^2 = .003$, and fear, $F(2, 77) = 1.20, p = .308, \eta^2 = .03$, 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 $\Delta$**

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 ($\phi$)

Used when both $X$ and $Y$ are dichotomous (i.e., both variables and both outcomes can be arranged on a $2 \times 2$ contingency table).

Example From the Literature

"The reliability of end-of-rotation evaluations was adequate (fellows, phi coefficient $[\phi] = 0.68$; faculty [including programme directors], $\phi = 0.71$)” (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 $\Delta$). 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 ($\alpha$), and critical effect size ($\Delta$). 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 ($\alpha$) at .05 (one-tailed), the power at .80, and the critical effect size ($\Delta$) 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

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>JJSP reading percentile</td>
<td>25.60</td>
<td>20.90</td>
<td>1–97</td>
<td>95</td>
</tr>
<tr>
<td>K-BIT vocabulary</td>
<td>76.92</td>
<td>17.69</td>
<td>40–111</td>
<td>95</td>
</tr>
<tr>
<td>K-BIT composite</td>
<td>80.57</td>
<td>16.22</td>
<td>48–121</td>
<td>95</td>
</tr>
<tr>
<td>PPVT-III</td>
<td>80.64</td>
<td>14.93</td>
<td>58–116</td>
<td>95</td>
</tr>
<tr>
<td>WRAT-3 reading subtest</td>
<td>85.57</td>
<td>18.66</td>
<td>45–117</td>
<td>95</td>
</tr>
</tbody>
</table>

**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**

\[ df = \text{Degrees of freedom} \]
\[ r_s = \text{Observed} \ r_s \text{ value} \]
\[ p = \text{Significance level} \]

**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**

\[ df = \text{Degrees of freedom} \]
\[ r = \text{Observed} \ r \text{ value} \]
\[ p = \text{Significance level} \]
\[ M = \text{Mean} \]
\[ SD = \text{Standard deviation} \]
### Table D3

Correlations Among Behavioral Variables and Offender Status: Male and Female ($n = 95$)

<table>
<thead>
<tr>
<th>Subscale</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Offender status</td>
<td>-</td>
<td>.03</td>
<td>-.18*</td>
<td>-.11</td>
<td>-.07</td>
<td>.15†</td>
<td>.04</td>
</tr>
<tr>
<td>2. Social stress</td>
<td>-</td>
<td>.46**</td>
<td>.70**</td>
<td>.49**</td>
<td>.43**</td>
<td>-.67</td>
<td></td>
</tr>
<tr>
<td>3. Inhibited</td>
<td>-</td>
<td>.31**</td>
<td>.30**</td>
<td>.42**</td>
<td>-.37**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Somatization</td>
<td>-</td>
<td>.47**</td>
<td>.44**</td>
<td>-.58**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Substance abuse</td>
<td>-</td>
<td></td>
<td>.29**</td>
<td>-.76**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Eating dysfunction</td>
<td>-</td>
<td></td>
<td></td>
<td>-.33**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Conforming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th>Condition</th>
<th>M (SD)</th>
<th>LL</th>
<th>UL</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-anxiety</td>
<td>.87 (.09)</td>
<td>.78</td>
<td>1.03</td>
</tr>
<tr>
<td>Low-anxiety</td>
<td>.68 (.10)</td>
<td>.54</td>
<td>.79</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Condition</th>
<th>Veterans n</th>
<th>M (SD)</th>
<th>95% CI</th>
<th>Rookies n</th>
<th>M (SD)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-anxiety</td>
<td>11</td>
<td>.81 (.07)</td>
<td>[.69, .89]</td>
<td>14</td>
<td>.90 (.14)</td>
<td>[.72, .94]</td>
</tr>
<tr>
<td>Low-anxiety</td>
<td>9</td>
<td>.65 (.11)</td>
<td>[.53, .76]</td>
<td>11</td>
<td>.70 (.13)</td>
<td>[.61, .82]</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. QOL</td>
<td>-</td>
<td>.38*</td>
<td>.64</td>
<td>58.3</td>
<td>8.2</td>
</tr>
<tr>
<td>2. BDI</td>
<td>.57*</td>
<td>-</td>
<td>.23</td>
<td>62.8</td>
<td>15.8</td>
</tr>
<tr>
<td>3. BAI</td>
<td>.34</td>
<td>.21</td>
<td>-</td>
<td>74.3</td>
<td>16.4</td>
</tr>
</tbody>
</table>

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.
Contrast Between Veterans and Rookies for Reaction Time in Low- and High-Anxiety Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>Veterans</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>t(45)</td>
<td>p</td>
<td>LL</td>
<td>UL</td>
</tr>
<tr>
<td>High-anxiety</td>
<td>.81</td>
<td>.07</td>
<td>.90</td>
<td>.14</td>
<td>1.45</td>
<td>.04</td>
<td>.73</td>
<td>.98</td>
</tr>
<tr>
<td>Low-anxiety</td>
<td>.65</td>
<td>.11</td>
<td>.70</td>
<td>.13</td>
<td>1.93</td>
<td>.12</td>
<td>.59</td>
<td>.76</td>
</tr>
</tbody>
</table>

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., $r_{CP1,CE1} = -0.544$), although when factoring in CE with a previous measure of CE and a performance time the relationship weakened (e.g., $r_{CE2,CE1*CP1} = -0.071; r_{CE2,CE1*T1} = 0.349$), 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

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Sample size</td>
</tr>
<tr>
<td>$N$</td>
<td>Population or lot size</td>
</tr>
<tr>
<td>$M$</td>
<td>Mean</td>
</tr>
<tr>
<td>$SD$</td>
<td>Standard deviation of the sample</td>
</tr>
<tr>
<td>$X$</td>
<td>Arithmetic mean of the sample</td>
</tr>
<tr>
<td>$s$</td>
<td>Standard deviation of the population</td>
</tr>
<tr>
<td>$CV$</td>
<td>Coefficient of variation</td>
</tr>
<tr>
<td>$s^2$</td>
<td>Sample variance</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>Variance of the population</td>
</tr>
<tr>
<td>$SE$</td>
<td>Standard error</td>
</tr>
<tr>
<td>$t$</td>
<td>$t$ distribution: shows both the random variable and a particular or observed value of this variable</td>
</tr>
<tr>
<td>$F$</td>
<td>$F$ distribution: shows both the random variable and a particular or observed value of this variable</td>
</tr>
<tr>
<td>$df$</td>
<td>Degree(s) of freedom</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>Chi-squared distribution: shows the random variable and a particular or observed value of this variable</td>
</tr>
<tr>
<td>$P$</td>
<td>Level of significance, probability</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Regression coefficient of population</td>
</tr>
<tr>
<td>$r$</td>
<td>Coefficient of correlation, sample</td>
</tr>
<tr>
<td>$r^2$</td>
<td>Coefficient of determination for $r$</td>
</tr>
<tr>
<td>$R$</td>
<td>Coefficient of multiple correlation</td>
</tr>
<tr>
<td>$R^2$</td>
<td>Coefficient of determination for $R$</td>
</tr>
<tr>
<td>$Cl$</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Gamma</td>
</tr>
<tr>
<td>$f$</td>
<td>Frequency</td>
</tr>
<tr>
<td>$\tau_b$</td>
<td>Kendall's tau-b</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Level of significance</td>
</tr>
<tr>
<td>Symbol</td>
<td>Explanation</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>µ</td>
<td>Population mean score</td>
</tr>
<tr>
<td>σ²</td>
<td>Population variance</td>
</tr>
<tr>
<td>ρ</td>
<td>Rho</td>
</tr>
<tr>
<td>z</td>
<td>z score</td>
</tr>
<tr>
<td>Z</td>
<td>Standardized scores, and Z distribution</td>
</tr>
<tr>
<td>r</td>
<td>Correlation</td>
</tr>
<tr>
<td>r_pb</td>
<td>Point biserial correlation</td>
</tr>
<tr>
<td>p</td>
<td>Statistical significance</td>
</tr>
<tr>
<td>Sig.</td>
<td>Statistical significance</td>
</tr>
<tr>
<td>95% CI</td>
<td>The 95% confidence interval</td>
</tr>
<tr>
<td>t</td>
<td>t test</td>
</tr>
<tr>
<td>F</td>
<td>One-way ANOVA or comparison test of between-group differences</td>
</tr>
<tr>
<td>η</td>
<td>Correlation coefficient for ANOVA (3+ categories)</td>
</tr>
<tr>
<td>Tukey-b</td>
<td>ANOVA test of significance of difference among means</td>
</tr>
<tr>
<td>Beta</td>
<td>Standardized regression slope</td>
</tr>
<tr>
<td>R²_change</td>
<td>Whether next variables entered add anything above variables entered</td>
</tr>
<tr>
<td>σ²_x</td>
<td>Variance</td>
</tr>
<tr>
<td>σ_xy</td>
<td>Covariance (of x and y)</td>
</tr>
</tbody>
</table>
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).

   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)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁</td>
<td>X</td>
<td>O₂</td>
</tr>
<tr>
<td>2</td>
<td>O₁</td>
<td>–</td>
<td>O₂</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ( (n = 50) )</td>
<td>R</td>
<td>BDI</td>
<td>Behavioral Therapy</td>
<td>BDI</td>
</tr>
<tr>
<td>2 ( (n = 50) )</td>
<td>R</td>
<td>BDI</td>
<td>-</td>
<td>BDI</td>
</tr>
</tbody>
</table>

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**.

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

<table>
<thead>
<tr>
<th>Test</th>
<th>df1</th>
<th>df2</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levene Statistic</td>
<td>.221</td>
<td>.98</td>
<td>.638</td>
</tr>
</tbody>
</table>

The Levene statistic is not significantly indicating homogeneity of variances.

Descriptive Statistics

<table>
<thead>
<tr>
<th>group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment group</td>
<td>50</td>
<td>31.700</td>
<td>6.12206</td>
<td>.98579</td>
<td>29.8601</td>
<td>33.4901</td>
<td>18.00</td>
<td>45.00</td>
</tr>
<tr>
<td>control group</td>
<td>50</td>
<td>33.720</td>
<td>6.92653</td>
<td>1.25946</td>
<td>32.0162</td>
<td>35.4348</td>
<td>22.80</td>
<td>48.00</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>32.710</td>
<td>6.98909</td>
<td>1.08008</td>
<td>31.5834</td>
<td>33.9188</td>
<td>18.00</td>
<td>49.00</td>
</tr>
</tbody>
</table>

Means and standard deviations illustrating the differences pre-to posttest.

Partial Eta Squared

- $\eta^2 = .01$ — small effect
- $\eta^2 = .09$ — medium effect
- $\eta^2 = .25$ — large effect

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.

Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type II Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
<th>Partial Eta Squared</th>
<th>Noncent. Parameter</th>
<th>Observed Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>763.565</td>
<td>7</td>
<td>109.329</td>
<td>.575</td>
<td>.002</td>
<td>.122</td>
<td>13.570</td>
<td>.670</td>
</tr>
<tr>
<td>Intercept</td>
<td>2678.215</td>
<td>1</td>
<td>2678.215</td>
<td>48.117</td>
<td>.000</td>
<td>.322</td>
<td>48.117</td>
<td>1.000</td>
</tr>
<tr>
<td>pretest</td>
<td>22.224</td>
<td>1</td>
<td>22.224</td>
<td>.383</td>
<td>.004</td>
<td>.004</td>
<td>.383</td>
<td>.004</td>
</tr>
<tr>
<td>group</td>
<td>763.982</td>
<td>1</td>
<td>763.982</td>
<td>13.500</td>
<td>.000</td>
<td>.133</td>
<td>13.500</td>
<td>.675</td>
</tr>
<tr>
<td>Error</td>
<td>5833.255</td>
<td>97</td>
<td>60.675</td>
<td>.000</td>
<td>.507</td>
<td>.000</td>
<td>.507</td>
<td>.050</td>
</tr>
<tr>
<td>Total</td>
<td>75376.000</td>
<td>100</td>
<td>75376.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Corrected Total | 6417.240 | 99 |

a. R Squared = .122 (Adjusted R Squared = .104)
b. Computed using alpha = .05

d. This parameter is set to zero because it is redundant.

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.

Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>t</th>
<th>Sig</th>
<th>95% Confidence Interval</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>31.695</td>
<td>4.440</td>
<td>7.135</td>
<td>.000</td>
<td>22.072</td>
<td>40.498</td>
</tr>
<tr>
<td>pretest</td>
<td>.679</td>
<td>.129</td>
<td>-.619</td>
<td>.538</td>
<td>-.333</td>
<td>.175</td>
</tr>
<tr>
<td>[group=1]</td>
<td>.568</td>
<td>1.548</td>
<td>-.3674</td>
<td>.000</td>
<td>-.874</td>
<td>2.812</td>
</tr>
<tr>
<td>[group=2]</td>
<td>0*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

<p>| Pre- and Posttest Control Group Design: Validity, Control, and Causal Inferences |
|---------------------------------|------------------|
| <strong>Internal Validity</strong>—Threat is (high ↑) (medium ↔) (low ↓) | <strong>Explanation</strong> |
| 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 |
| <strong>Statistical Conclusion Validity</strong>—Threat is (high ↑) (medium ↔) (low ↓) | <strong>Explanation</strong> |
| 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 |</p>
<table>
<thead>
<tr>
<th>Restriction of range</th>
<th>↓</th>
<th>Treatment and control conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraneous variance in the experimental setting</td>
<td>↔</td>
<td>Strong levels of design control implemented</td>
</tr>
<tr>
<td>Inaccurate effect size estimation</td>
<td>↓</td>
<td>One-way ANCOVA is the most appropriate analysis</td>
</tr>
<tr>
<td>Variability in procedures</td>
<td>↓</td>
<td>Strong levels of design control implemented</td>
</tr>
<tr>
<td>Subject heterogeneity</td>
<td>↓</td>
<td>Random selection and random assignment</td>
</tr>
<tr>
<td>Unreliability of the measures</td>
<td>↓</td>
<td>BDI maintains adequate psychometric properties</td>
</tr>
<tr>
<td>Multiple comparisons and error rates</td>
<td>↓</td>
<td>Single dependent variable and one primary analysis used</td>
</tr>
</tbody>
</table>

**Control—Design and statistical control is (strong ⌂) (medium ⌂) (weak ⌂)**

| Manipulation | ⌂ rtl | Established therapeutic approach w/ manipulation check |
| Elimination | ⌂ rtl | Established therapeutic approach w/ manipulation check |
| Inclusion | ⌂ rtl | Multiple groups compared (included treatment and control) |
| Group or condition assignment | ⌂ rtl | Random assignment |
| Statistical procedures | rtl | Pretest entered as covariates |

**Cause and Effect—Cause-effect determination is (strong ⌂ rtl) (medium ⌂) (weak ⌂)**

| Covariation | rtl | Strong controls and minimal threats to internal and statistical conclusion validity |
| Temporal precedence | rtl | Strong controls and minimal threats to internal and statistical conclusion validity |
| No plausible alternative explanations | rtl | 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)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>O₁</td>
<td>Xₐ</td>
<td>O₂</td>
</tr>
<tr>
<td>2</td>
<td>O₁</td>
<td>X₉</td>
<td>O₂</td>
</tr>
</tbody>
</table>

**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, \( H₀: \mu₁ = \mu₂ \) 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.

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n = 40)</td>
<td>R</td>
<td>EQ-i</td>
<td>In-class social skills training</td>
<td>EQ-i</td>
</tr>
<tr>
<td>2 (n = 40)</td>
<td>R</td>
<td>EQ-i</td>
<td>Parent-centered EQ training</td>
<td>EQ-i</td>
</tr>
</tbody>
</table>

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 \( rpb \), 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:

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSA</td>
<td>40</td>
<td>14.3020</td>
<td>8.66145</td>
<td>1.27261</td>
</tr>
<tr>
<td>TreatmentA</td>
<td>40</td>
<td>8.5725</td>
<td>11.22928</td>
<td>1.67551</td>
</tr>
</tbody>
</table>

Treatment A on average scored 5.72 higher than Treatment B.

The \( t \) statistic shows significance \( (p < .05) \).

Mean scores on the EQI between the pre- and posttests for both treatment groups:

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>pretest</th>
<th>posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>TreatmentA</td>
<td>40</td>
<td>118.8930</td>
<td>116.8030</td>
</tr>
<tr>
<td>TreatmentB</td>
<td>40</td>
<td>104.3000</td>
<td>102.8972</td>
</tr>
</tbody>
</table>

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, 2008, 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' $g$.

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

$$d = \frac{14.3 - 8.575}{\sqrt{\frac{(39)(9.661)^2 + (39)(11.229)^2}{40 + 40 - 2}}} = \frac{5.725}{10.474} = 0.5466 = .55$$

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.
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 \((r_{pb} = -.27; R^2 = -.071)\).
(Continued)

<table>
<thead>
<tr>
<th>Threat to Validity</th>
<th>Strength of Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraneous variance in the experimental setting</td>
<td>↔</td>
</tr>
<tr>
<td>Inaccurate effect size estimation</td>
<td>↔</td>
</tr>
<tr>
<td>Variability in procedures</td>
<td>↓</td>
</tr>
<tr>
<td>Subject heterogeneity</td>
<td>↓</td>
</tr>
<tr>
<td>Unreliability of the measures</td>
<td>↓</td>
</tr>
<tr>
<td>Multiple comparisons and error rates</td>
<td>↓</td>
</tr>
</tbody>
</table>

**Control**—Design and statistical control is (strong ☑️) (medium ☐️) (weak ❌)

<table>
<thead>
<tr>
<th>Threat to Validity</th>
<th>Strength of Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manipulation</td>
<td>☑️</td>
</tr>
<tr>
<td>Elimination</td>
<td>☑️</td>
</tr>
<tr>
<td>Inclusion</td>
<td>☑️</td>
</tr>
<tr>
<td>Group or condition assignment</td>
<td>☑️</td>
</tr>
<tr>
<td>Statistical procedures</td>
<td>☑️</td>
</tr>
</tbody>
</table>

**Cause and Effect**—Cause-effect determination is (strong ☑️) (medium ☐️) (weak ❌)

<table>
<thead>
<tr>
<th>Threat to Validity</th>
<th>Strength of Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariation</td>
<td>☑️</td>
</tr>
<tr>
<td>Temporal precedence</td>
<td>☑️</td>
</tr>
<tr>
<td>No plausible alternative explanations</td>
<td>☑️</td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest1</th>
<th>Posttest2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0₁</td>
<td>Xₐ</td>
<td>0₂</td>
<td>0₃</td>
</tr>
<tr>
<td>2</td>
<td>0₁</td>
<td>X₈</td>
<td>0₂</td>
<td>0₃</td>
</tr>
</tbody>
</table>

**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, H₀: μ₁ ≠ μ₂ and μ₃ = μ₄ 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.

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Pretest</th>
<th>Treatment</th>
<th>Posttest1</th>
<th>Posttest2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n = 25)</td>
<td>R</td>
<td>QOLFS</td>
<td>BFB therapy</td>
<td>QOLFS</td>
<td>QOLFS</td>
</tr>
<tr>
<td>2 (n = 25)</td>
<td>R</td>
<td>QOLFS</td>
<td>Yoga</td>
<td>QOLFS</td>
<td>QOLFS</td>
</tr>
</tbody>
</table>

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
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.

Enter each time point in order in the WS Variables box. Add group to the BS Factors box, then select Plots, and then Options.
Select Add and then Continue.
Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre</td>
<td>3.8800</td>
<td>1.26886</td>
<td>25</td>
</tr>
<tr>
<td>Yoga</td>
<td>4.0400</td>
<td>.97809</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>3.9600</td>
<td>1.12413</td>
<td>50</td>
</tr>
<tr>
<td>post1</td>
<td>6.8800</td>
<td>.97125</td>
<td>25</td>
</tr>
<tr>
<td>Yoga</td>
<td>4.9200</td>
<td>1.11505</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>5.9000</td>
<td>1.43214</td>
<td>50</td>
</tr>
<tr>
<td>post2</td>
<td>7.1200</td>
<td>.97125</td>
<td>25</td>
</tr>
<tr>
<td>Yoga</td>
<td>5.6000</td>
<td>1.08012</td>
<td>25</td>
</tr>
<tr>
<td>Total</td>
<td>6.3600</td>
<td>1.27391</td>
<td>50</td>
</tr>
</tbody>
</table>

Box's Test of Equality of Covariance Matrices

<table>
<thead>
<tr>
<th></th>
<th>Box's M</th>
<th>F</th>
<th>df</th>
<th>df2</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>pre</td>
<td>3.401</td>
<td>5.28</td>
<td>6</td>
<td>16.993</td>
<td>.122</td>
</tr>
<tr>
<td>post1</td>
<td>5.420</td>
<td>6.00</td>
<td>6</td>
<td>16.993</td>
<td>.122</td>
</tr>
</tbody>
</table>

Box's M test is not significant, indicating multivariate normality.

Mauchly's Test of Sphericity

<table>
<thead>
<tr>
<th>Source</th>
<th>Approx. Chi-Square</th>
<th>df</th>
<th>Sig.</th>
<th>Deflation-Not Deflation</th>
<th>Full</th>
<th>Lower-bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time_Point</td>
<td>182.250</td>
<td>2</td>
<td>0.000</td>
<td>66.200</td>
<td>66.200</td>
<td>66.200</td>
</tr>
<tr>
<td>Time_Point * group</td>
<td>31.250</td>
<td>2</td>
<td>0.000</td>
<td>15.647</td>
<td>15.647</td>
<td>15.647</td>
</tr>
</tbody>
</table>

Mauchly's test is not significant, indicating sphericity.

Mauchly's test is not significant; therefore, the "Sphericity Assumed" row should be used.

Tests of Within-Subjects Contrasts

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time_Point</td>
<td>144.390</td>
<td>1</td>
<td>144.390</td>
<td>118.431</td>
<td>.001</td>
<td>.712</td>
</tr>
<tr>
<td>Time_Point * group</td>
<td>18.353</td>
<td>1</td>
<td>18.353</td>
<td>13.960</td>
<td>.001</td>
<td>.225</td>
</tr>
<tr>
<td>Error(Time_Point)</td>
<td>17.840</td>
<td>1</td>
<td>17.840</td>
<td>13.960</td>
<td>.001</td>
<td>.225</td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4,084.607</td>
<td>1</td>
<td>4,084.607</td>
<td>492.310</td>
<td>.000</td>
<td>.990</td>
</tr>
<tr>
<td>group</td>
<td>45.927</td>
<td>1</td>
<td>45.927</td>
<td>50.561</td>
<td>.000</td>
<td>.513</td>
</tr>
<tr>
<td>Error</td>
<td>43.602</td>
<td>48</td>
<td>.909</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PAIRWISE COMPARISONS

Data $\rightarrow$ Split File
Note: Discard all other output except for the Pairwise Comparisons.

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 x 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 x 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 ↑) (medium ↔) (low ↓) |
| History | ↑ | Inclusion of pretest |
| Maturation | ↑ | Inclusion of pretest |
| Testing | ↑ | Inclusion of pretest |
| Instrumentation | ↓ | The instrument did not change between pre- and posttests |
| Statistical regression | ↔ | Random assignment applied |
| Attrition | ↔ | Length of time between measures |
| 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 | Each group was exposed to only one condition |
| **Statistical Conclusion Validity**—Threat is (high ↑) (medium ↔) (low ↓) |
| Low statistical power | ↓ | Adequate sample size |
| Assumption violation of statistical tests | ↓ | Multivariate normality, sphericity, and independence confirmed |
Appendix E 277

2-Factor Pre- and Double Posttest Design: Validity, Control, and Causal Inferences

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error rate problem</td>
<td>↓ Single primary analysis used</td>
</tr>
<tr>
<td>Restriction of range</td>
<td>↓ Distinct treatment conditions</td>
</tr>
<tr>
<td>Extraneous variance in the experimental setting</td>
<td>↓ Strong levels of design control implemented</td>
</tr>
<tr>
<td>Inaccurate effect size estimation</td>
<td>↓ RM ANOVA is the most appropriate analysis for this design</td>
</tr>
<tr>
<td>Variability in procedures</td>
<td>↓ Strong levels of design control implemented</td>
</tr>
<tr>
<td>Subject heterogeneity</td>
<td>↓ Random assignment</td>
</tr>
<tr>
<td>Unreliability of the measures</td>
<td>↓ QOLFS maintains adequate psychometric properties</td>
</tr>
<tr>
<td>Multiple comparisons and error rates</td>
<td>↓ Single dependent variable and one primary analysis used</td>
</tr>
</tbody>
</table>

Control—Design and statistical control is (strong ⚫⚫⚫) (medium ⚫⚫) (weak ⚫)

| ⚫⚫⚫ | Established programs w/ manipulation check |
| ⚫⚫⚫ | Established programs w/ manipulation check |
| ⚫⚫⚫ | Multiple groups compared (two treatment conditions) |
| ⚫⚫⚫ | Random assignment |
| n/a | Data were not altered |

Cause and Effect—Cause–effect determination is (strong ⚫⚫⚫) (medium ⚫⚫) (weak ⚫)

| ⚫⚫⚫ | Strong controls and minimal threats to internal and statistical conclusion validity |
| ⚫⚫⚫ | Strong controls and minimal threats to internal and statistical conclusion validity |
| ⚫⚫ | 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: $F(1, 42) = 50.56, p = .000$. 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 ($M = 4.92$, $SD = .97$).
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 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)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Condition</th>
<th>Test</th>
<th>Condition</th>
<th>Test</th>
<th>Condition</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X_A</td>
<td>O_1</td>
<td>X_B</td>
<td>O_2</td>
<td>X_C</td>
<td>O_3</td>
</tr>
<tr>
<td>2</td>
<td>X_B</td>
<td>O_1</td>
<td>X_C</td>
<td>O_2</td>
<td>X_A</td>
<td>O_3</td>
</tr>
<tr>
<td>3</td>
<td>X_C</td>
<td>O_1</td>
<td>X_A</td>
<td>O_2</td>
<td>X_B</td>
<td>O_3</td>
</tr>
<tr>
<td>4</td>
<td>X_C</td>
<td>O_1</td>
<td>X_B</td>
<td>O_2</td>
<td>X_A</td>
<td>O_3</td>
</tr>
<tr>
<td>5</td>
<td>X_A</td>
<td>O_1</td>
<td>X_C</td>
<td>O_2</td>
<td>X_B</td>
<td>O_3</td>
</tr>
<tr>
<td>6</td>
<td>X_B</td>
<td>O_1</td>
<td>X_A</td>
<td>O_2</td>
<td>X_C</td>
<td>O_3</td>
</tr>
</tbody>
</table>

*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 \( \mu_3 \) 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.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Condition</th>
<th>Test</th>
<th>Condition</th>
<th>Test</th>
<th>Condition</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>RT</td>
<td>Medium</td>
<td>RT</td>
<td>Low</td>
<td>RT</td>
</tr>
<tr>
<td>2</td>
<td>Medium</td>
<td>RT</td>
<td>Low</td>
<td>RT</td>
<td>High</td>
<td>RT</td>
</tr>
<tr>
<td>3</td>
<td>Low</td>
<td>RT</td>
<td>High</td>
<td>RT</td>
<td>Medium</td>
<td>RT</td>
</tr>
<tr>
<td>4</td>
<td>Low</td>
<td>RT</td>
<td>Medium</td>
<td>RT</td>
<td>High</td>
<td>RT</td>
</tr>
<tr>
<td>5</td>
<td>High</td>
<td>RT</td>
<td>Low</td>
<td>RT</td>
<td>Medium</td>
<td>RT</td>
</tr>
<tr>
<td>6</td>
<td>Medium</td>
<td>RT</td>
<td>High</td>
<td>RT</td>
<td>Low</td>
<td>RT</td>
</tr>
</tbody>
</table>

*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.
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 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 ($M = 406.50$, $SD = 13.61$).
### Appendix E

#### 3-Factor Crossover Design: Validity, Control, and Causal Inferences

<table>
<thead>
<tr>
<th><strong>Internal Validity</strong>—Threat is (high ↑) (medium ↔) (low ↓)</th>
<th><strong>Level</strong></th>
<th><strong>Explanation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing</td>
<td>↑</td>
<td>Multiple measures taken over time</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>↓</td>
<td>Measures stayed the same throughout</td>
</tr>
<tr>
<td>Statistical regression</td>
<td>↔</td>
<td>Random assignment applied</td>
</tr>
<tr>
<td>Attrition</td>
<td>↔</td>
<td>Length of time between measures</td>
</tr>
<tr>
<td>Selection bias</td>
<td>↓</td>
<td>Conditions were randomly counterbalanced</td>
</tr>
<tr>
<td>Combination of selection and other treatments</td>
<td>↔</td>
<td>Each subject was exposed to all three conditions</td>
</tr>
<tr>
<td>Diffusion</td>
<td>↔</td>
<td>Participation in one condition can affect the other</td>
</tr>
<tr>
<td>Special treatment</td>
<td>↓</td>
<td>All conditions were administered following the protocols</td>
</tr>
<tr>
<td>Sequencing effects</td>
<td>↔</td>
<td>Conditions were counterbalanced</td>
</tr>
</tbody>
</table>

#### Statistical Conclusion Validity—Threat is (high ↑) (medium ↔) (low ↓)

<table>
<thead>
<tr>
<th><strong>Statistical Conclusion Validity</strong>—Threat is (high ↑) (medium ↔) (low ↓)</th>
<th><strong>Level</strong></th>
<th><strong>Explanation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low statistical power</td>
<td>↔</td>
<td>Small sample size but multiple measures taken over time</td>
</tr>
<tr>
<td>Assumption violation of statistical tests</td>
<td>↓</td>
<td>Sphericity and independence confirmed</td>
</tr>
<tr>
<td>Error rate problem</td>
<td>↓</td>
<td>Single primary analysis used</td>
</tr>
<tr>
<td>Restriction of range</td>
<td>↓</td>
<td>Distinct treatment conditions</td>
</tr>
<tr>
<td>Extraneous variance in the experimental setting</td>
<td>↓</td>
<td>Strong levels of design control implemented</td>
</tr>
<tr>
<td>Inaccurate effect size estimation</td>
<td>↓</td>
<td>RM ANOVA is the most appropriate analysis for this design</td>
</tr>
<tr>
<td>Variability in procedures</td>
<td>↓</td>
<td>Strong levels of design control implemented</td>
</tr>
<tr>
<td>Subject heterogeneity</td>
<td>↔</td>
<td>Subjects share likenesses</td>
</tr>
<tr>
<td>Unreliability of the measures</td>
<td>↓</td>
<td>Simple reaction time measured electronically w/o error</td>
</tr>
<tr>
<td>Multiple comparisons and error rates</td>
<td>↓</td>
<td>Single dependent variable and one primary analysis used</td>
</tr>
</tbody>
</table>

#### Control—Design and statistical control is (strong ⬤) (medium ⬤) (weak ↑)

<table>
<thead>
<tr>
<th><strong>Control</strong>—Design and statistical control is (strong ⬤) (medium ⬤) (weak ↑)</th>
<th><strong>Level</strong></th>
<th><strong>Explanation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manipulation</td>
<td>⬤ ⬤ ⬤</td>
<td>Established protocol w/ manipulation check</td>
</tr>
<tr>
<td>Elimination</td>
<td>⬤ ⬤ ⬤</td>
<td>Established protocol w/ manipulation check</td>
</tr>
<tr>
<td>Inclusion</td>
<td>⬤ ⬤ ⬤</td>
<td>Three conditions included</td>
</tr>
<tr>
<td>Group or condition assignment</td>
<td>⬤ ⬤ ⬤</td>
<td>Counterbalanced conditions</td>
</tr>
<tr>
<td>Statistical procedures</td>
<td>n/a</td>
<td>Data were not altered</td>
</tr>
</tbody>
</table>

(Continued)
3-Factor Crossover Design: Validity, Control, and Causal Inferences

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause and Effect—Cause-effect determination is (strong ⬤) (medium ⬤) (weak ⬤)</td>
<td></td>
</tr>
<tr>
<td>Covariation</td>
<td>⬤ ⬤</td>
</tr>
<tr>
<td>Temporal precedence</td>
<td>⬤ ⬤</td>
</tr>
<tr>
<td>No plausible alternative explanations</td>
<td>⬤ ⬤</td>
</tr>
</tbody>
</table>

**Example E5 3 x 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 x 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 x 3 Latin-Square Design

<table>
<thead>
<tr>
<th>Group</th>
<th>Blocking Factor 1</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Level 1</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>Level 2</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>Level 3</td>
<td>B</td>
<td>C</td>
<td>A</td>
</tr>
</tbody>
</table>

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 x 3 Latin-square design

**Null Hypothesis:** The adjusted population means for all groups are equal. In this example, $H_0: t_1 = t_2 = \ldots = t_p = 0$ or $H_0: \bar{T}_1 = 0$.

**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, $3_1 * 3_2 = 9$)
Putt Difficulty

<table>
<thead>
<tr>
<th>Group</th>
<th>Anxiety</th>
<th>Easy</th>
<th>Medium</th>
<th>Hard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n = 9)</td>
<td>High</td>
<td>A (Blade)</td>
<td>B (Mallet)</td>
<td>C (Futuristic)</td>
</tr>
<tr>
<td>2 (n = 9)</td>
<td>Medium</td>
<td>C (Futuristic)</td>
<td>A (Blade)</td>
<td>B (Mallet)</td>
</tr>
<tr>
<td>3 (n = 9)</td>
<td>Low</td>
<td>B (Mallet)</td>
<td>C (Futuristic)</td>
<td>A (Blade)</td>
</tr>
</tbody>
</table>

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 $3 \times 3 \times 3$ 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**

<table>
<thead>
<tr>
<th>Source</th>
<th>Type I Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>6742.815*</td>
<td>6</td>
<td>1457.136</td>
<td>333.846</td>
<td>.000</td>
<td>.564</td>
</tr>
<tr>
<td>Intercept</td>
<td>13585.180</td>
<td>1</td>
<td>13585.180</td>
<td>312.517</td>
<td>.000</td>
<td>.577</td>
</tr>
<tr>
<td>Anxiety</td>
<td>1439.210</td>
<td>2</td>
<td>719.605</td>
<td>184.869</td>
<td>.000</td>
<td>.517</td>
</tr>
<tr>
<td>Difficulty</td>
<td>6623.877</td>
<td>2</td>
<td>3311.938</td>
<td>793.188</td>
<td>.000</td>
<td>.585</td>
</tr>
<tr>
<td>Putter_Type</td>
<td>379.278</td>
<td>2</td>
<td>199.684</td>
<td>43.500</td>
<td>.000</td>
<td>.540</td>
</tr>
<tr>
<td>Error</td>
<td>322.808</td>
<td>74</td>
<td>4.365</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23651.000</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>90265.900</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ R^2 = .564 \text{ (Adjusted } R^2 = .551) \]

### Multiple Comparisons

**Dependent Variable: Performance**

<table>
<thead>
<tr>
<th>Tukey HSD</th>
<th>Mean Difference (J)</th>
<th>Std. Error</th>
<th>Sig</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Anxiety (B) Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High Medium</td>
<td>9.3704</td>
<td>.56800</td>
<td>.000</td>
<td>5.0104 - 7.3303</td>
</tr>
<tr>
<td>Medium High</td>
<td>-10.2224</td>
<td>.56800</td>
<td>.000</td>
<td>-15.5822 - -4.8623</td>
</tr>
<tr>
<td>Low Medium High</td>
<td>-10.2224</td>
<td>.56800</td>
<td>.000</td>
<td>-15.5822 - -4.8623</td>
</tr>
<tr>
<td>Medium Low</td>
<td>-9.5185</td>
<td>.56800</td>
<td>.000</td>
<td>-10.8785 - -8.1581</td>
</tr>
<tr>
<td>High Medium Low</td>
<td>-9.5185</td>
<td>.56800</td>
<td>.000</td>
<td>-10.8785 - -8.1581</td>
</tr>
</tbody>
</table>

### Performance

<table>
<thead>
<tr>
<th>Tukey B a,b</th>
<th>Mean Difference (J)</th>
<th>Std. Error</th>
<th>Sig</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Difficulty (B) Difficulty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy Medium</td>
<td>-22.6255</td>
<td>.56800</td>
<td>.000</td>
<td>-23.9155 - -21.3355</td>
</tr>
<tr>
<td>Medium Easy</td>
<td>9.5185</td>
<td>.56800</td>
<td>.000</td>
<td>8.1581 - 10.8785</td>
</tr>
<tr>
<td>Hard Easy</td>
<td>-13.0376</td>
<td>.56800</td>
<td>.000</td>
<td>-14.3970 - -11.6771</td>
</tr>
<tr>
<td>Medium Medium</td>
<td>13.0376</td>
<td>.56800</td>
<td>.000</td>
<td>11.6771 - 14.3970</td>
</tr>
<tr>
<td>Medium Hard</td>
<td>11.7778</td>
<td>.56800</td>
<td>.000</td>
<td>11.6771 - 14.3970</td>
</tr>
</tbody>
</table>

**Tukey HSD**

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>N</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>27</td>
<td>2.2593</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>27</td>
<td>11.7778</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hard</td>
<td>27</td>
<td></td>
<td></td>
<td>24.8148</td>
</tr>
</tbody>
</table>

**Dependent Variable: Performance**

<table>
<thead>
<tr>
<th>Tukey B a,b</th>
<th>Mean Difference (J)</th>
<th>Std. Error</th>
<th>Sig</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Putter_Type (B) Putter_Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blade Mallet</td>
<td>-4.4148</td>
<td>.56800</td>
<td>.000</td>
<td>-5.7148 - -3.1148</td>
</tr>
<tr>
<td>Blade Mallet Futuristic</td>
<td>-4.4148</td>
<td>.56800</td>
<td>.000</td>
<td>-5.7148 - -3.1148</td>
</tr>
<tr>
<td>Mallet Blade</td>
<td>4.8148</td>
<td>.56800</td>
<td>.000</td>
<td>3.4548 - 6.1748</td>
</tr>
<tr>
<td>Mallet Blade Futuristic</td>
<td>4.8148</td>
<td>.56800</td>
<td>.000</td>
<td>3.4548 - 6.1748</td>
</tr>
</tbody>
</table>

**Performance**

<table>
<thead>
<tr>
<th>Puller Type</th>
<th>N</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blade</td>
<td>27</td>
<td>11.1852</td>
<td></td>
</tr>
<tr>
<td>Futuristic</td>
<td>27</td>
<td>11.6657</td>
<td></td>
</tr>
<tr>
<td>Mallet</td>
<td>27</td>
<td>16.0000</td>
<td></td>
</tr>
</tbody>
</table>

*Differences between the blade and futuristic were not significant.*
Estimated Marginal Means of Performance at Difficulty = Hard

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Anxiety</th>
<th>Putter Type</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>High</td>
<td>Blade</td>
<td>3.5556</td>
<td>.88192</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>3.5556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>Futuristic</td>
<td>Total</td>
<td>1.6667</td>
<td>.70711</td>
</tr>
<tr>
<td>Low</td>
<td>Mallet</td>
<td>1.5556</td>
<td></td>
<td>.52705</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1.5556</td>
<td></td>
<td>.52705</td>
</tr>
<tr>
<td>Medium</td>
<td>High</td>
<td>Mallet</td>
<td>21.8889</td>
<td>1.16667</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>21.8889</td>
<td></td>
<td>1.16667</td>
</tr>
<tr>
<td>Low</td>
<td>Futuristic</td>
<td>Total</td>
<td>3.3333</td>
<td>.50000</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>3.3333</td>
<td></td>
<td>.50000</td>
</tr>
<tr>
<td>Hard</td>
<td>High</td>
<td>Futuristic</td>
<td>30.0000</td>
<td>1.41421</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>30.0000</td>
<td></td>
<td>1.41421</td>
</tr>
<tr>
<td>Medium</td>
<td>Mallet</td>
<td>24.5555</td>
<td></td>
<td>1.13039</td>
</tr>
<tr>
<td>Low</td>
<td>Blade</td>
<td>19.8889</td>
<td></td>
<td>1.45297</td>
</tr>
<tr>
<td>Latin-Square Design: Validity, Control, and Causal Inferences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level</strong></td>
<td><strong>Explanation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Internal Validity—Threat is (high ↑) (medium ↔) (low ↓)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>↑</td>
<td>Multiple measures taken over time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maturation</td>
<td>↑</td>
<td>Multiple measures taken over time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td>↑</td>
<td>Multiple measures taken over time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrumentation</td>
<td>↓</td>
<td>Measures stayed the same throughout</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical regression</td>
<td>↔</td>
<td>Random assignment applied</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attrition</td>
<td>↔</td>
<td>Length of time between measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection bias</td>
<td>↓</td>
<td>Conditions were randomly counterbalanced</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combination of selection and other treatments</td>
<td>↓</td>
<td>Each subject was exposed to all three conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diffusion</td>
<td>↔</td>
<td>Participation in one condition can affect the other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special treatment</td>
<td>↓</td>
<td>All conditions were administered following the protocols</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequencing effects</td>
<td>↑</td>
<td>Conditions were counterbalanced but may affect performance in the subsequent conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Statistical Conclusion Validity—Threat is (high ↑) (medium ↔) (low ↓)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low statistical power</td>
<td>↓</td>
<td>Adequate sample size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assumption violation of statistical tests</td>
<td>↔</td>
<td>Data were normally distributed, but interaction effects may be present between variables but were not tested</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error rate problem</td>
<td>↓</td>
<td>Single primary analysis used</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restriction of range</td>
<td>↓</td>
<td>Distinct treatment conditions resulted in wide ranges in performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraneous variance in the experimental setting</td>
<td>↓</td>
<td>Strong levels of design control implemented</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inaccurate effect size estimation</td>
<td>↓</td>
<td>ANOVA is the most appropriate analysis for this design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variability in procedures</td>
<td>↓</td>
<td>Strong levels of design control implemented</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject heterogeneity</td>
<td>↔</td>
<td>Subjects share likenesses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unreliability of the measures</td>
<td>↓</td>
<td>Radial distance measured in centimeters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple comparisons and error rates</td>
<td>↓</td>
<td>Single dependent variable and one primary analysis used</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
Latin-Square Design: Validity, Control, and Causal Inferences

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control—Design and statistical control is (strong †††) (medium ††) (weak †)</td>
<td></td>
</tr>
<tr>
<td>Manipulation</td>
<td>††† Established protocol w/ manipulation check</td>
</tr>
<tr>
<td>Elimination</td>
<td>††† Established protocol w/ manipulation check</td>
</tr>
<tr>
<td>Inclusion</td>
<td>††† Multiple conditions included</td>
</tr>
<tr>
<td>Group or condition assignment</td>
<td>††† Counterbalanced conditions</td>
</tr>
<tr>
<td>Statistical procedures</td>
<td>n/a Data were not altered</td>
</tr>
</tbody>
</table>

Cause and Effect—Cause–effect determination is (strong †††) (medium ††) (weak †)

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariation</td>
<td>†† Strong controls and medium threats to internal and statistical conclusion validity</td>
</tr>
<tr>
<td>Temporal precedence</td>
<td>†† Strong controls and medium threats to internal and statistical conclusion validity</td>
</tr>
<tr>
<td>No plausible alternative explanations</td>
<td>†† Covariation and temporal precedence confirmed, but replicable results should be obtained over time</td>
</tr>
</tbody>
</table>

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 (O₁ and O₂). 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 (O₁ and O₂) 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\(_3\)).

**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)

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Test</th>
<th>Treatment</th>
<th>Test</th>
<th>–</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(X_A)</td>
<td>(O_1)</td>
<td>–</td>
<td>(O_2)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Time delay</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>–</td>
<td>(O_3)</td>
<td>(X_A)</td>
<td>(O_4)</td>
<td>–</td>
<td>(O_5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Time delay</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*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 \(O_5\); 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.

<table>
<thead>
<tr>
<th>Group</th>
<th>Assignment</th>
<th>Treatment</th>
<th>Test1</th>
<th>Treatment</th>
<th>Test2</th>
<th>–</th>
<th>Test3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n = 16)</td>
<td>R</td>
<td>Omega-3 oil</td>
<td>LDL</td>
<td>Time delay</td>
<td>LDL</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2 (n = 16)</td>
<td>R</td>
<td>Placebo</td>
<td>LDL</td>
<td>Omega-3 oil</td>
<td>LDL</td>
<td>Time delay</td>
<td>LDL</td>
</tr>
</tbody>
</table>

Sample Data Collection Spreadsheet

<table>
<thead>
<tr>
<th>Participant</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>207</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>182</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>272</td>
<td>19</td>
</tr>
<tr>
<td>16</td>
<td>210</td>
<td>32</td>
</tr>
<tr>
<td>16</td>
<td>246</td>
<td>247</td>
</tr>
<tr>
<td>32</td>
<td>192</td>
<td>261</td>
</tr>
</tbody>
</table>

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.
## COMMAND PROMPTS FOR SPSS

Data titled *waitlistcontinuation.sav*

**Plot the Data**

**Graphs → Legacy Dialogs → Boxplot → Simple**
Descriptives

Compare Means → Means
Mean gain scores between the two treatment groups

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
<th>Median</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>group 1</td>
<td>206.8125</td>
<td>16</td>
<td>33.96315</td>
<td>202.0000</td>
<td>-.102</td>
<td>.276</td>
</tr>
<tr>
<td>group 2</td>
<td>239.0000</td>
<td>16</td>
<td>29.61981</td>
<td>239.0000</td>
<td>-.473</td>
<td>-.102</td>
</tr>
<tr>
<td>Total</td>
<td>222.9063</td>
<td>32</td>
<td>35.35567</td>
<td>222.0000</td>
<td>-.525</td>
<td>-.072</td>
</tr>
</tbody>
</table>

Independent Samples t Test

Analyze → Compare Means → Independent Samples t Test

As a general rule, Skewness and Kurtosis within +1 to -1 indicates that the data is normally distributed.
The $t$ statistic shows significance ($p < .05$).

Group 1 (treatment group) has LDL levels 32.19 lower than Treatment B.

The Levene statistic is not significant, indicating homogeneity of variances.

Point Biserial Correlation Coefficient (Effect Size)

Graph → Legacy Dialogues → 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

<table>
<thead>
<tr>
<th></th>
<th>collapsedpost</th>
<th>collapsedfollowup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>209.4687</td>
<td>238.4375</td>
</tr>
<tr>
<td>N</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>28.33553</td>
<td>22.88955</td>
</tr>
<tr>
<td>Median</td>
<td>208.0000</td>
<td>241.0000</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>.001</td>
<td>.091</td>
</tr>
<tr>
<td>Skewness</td>
<td>.072</td>
<td>-.360</td>
</tr>
</tbody>
</table>

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 $\rightarrow$ Compare Means $\rightarrow$ Paired-Samples $t$ Test
### Effect Size

**Converting $t$-value into $r$-value**

$$r = \frac{\sqrt{t^2}}{t^2 + df}$$
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 O₂). 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., \( O_1 + O_4 \) compared to \( O_2 + O_5 \)) 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 (\( M = 209.47, SD = 28.33 \)).

<table>
<thead>
<tr>
<th>Wait-List Continuation Design: Validity, Control, and Causal Inferences</th>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal Validity</strong>—Threat is (high ↑) (medium ↔) (low ↓)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>↑</td>
<td>Inclusion of multiple tests, including a pretest for Group 2</td>
</tr>
<tr>
<td>Maturation</td>
<td>↑</td>
<td>Inclusion of multiple tests, including a pretest for Group 2</td>
</tr>
<tr>
<td>Testing</td>
<td>↑</td>
<td>Inclusion of multiple tests, including a pretest for Group 2</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>↓</td>
<td>Assessment of LDLS is stable</td>
</tr>
</tbody>
</table>

(Continued)
### Wait-List Continuation Design: Validity, Control, and Causal Inferences

<table>
<thead>
<tr>
<th></th>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal Validity</strong>—Threat is  (high ↑) (medium ↔) (low ↓)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical regression</td>
<td>↔</td>
<td>Random assignment applied and time delay</td>
</tr>
<tr>
<td>Attrition</td>
<td>↔</td>
<td>Length of time between measures, including time delay</td>
</tr>
<tr>
<td>Selection bias</td>
<td>↓</td>
<td>Random assignment applied</td>
</tr>
<tr>
<td>Combination of selection and other treatments</td>
<td>↓</td>
<td>Random assignment and independent groups</td>
</tr>
<tr>
<td>Diffusion</td>
<td>↓</td>
<td>Conditions are independent but both groups receive treatment</td>
</tr>
<tr>
<td>Special treatment</td>
<td>↓</td>
<td>Independent treatment conditions</td>
</tr>
<tr>
<td>Sequencing effects</td>
<td>↓</td>
<td>Conditions are independent but both groups receive treatment</td>
</tr>
</tbody>
</table>

| **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 | ↓     | $t$ 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 BBB) (medium BB) (weak B) |       |                                                 |
| Manipulation | BBB   | Controlled for patient diet and activity level   |
| Elimination  | BBB   | Controlled for patient diet and activity level   |
| Inclusion    | BBB   | Multiple groups compared (two treatment conditions) |
| Group or condition assignment | BBB | Random assignment                               |
| Statistical procedures | BB | Data collapsed to examine lasting effects        |
### Wait-List Continuation Design: Validity, Control, and Causal Inferences

<table>
<thead>
<tr>
<th><strong>Cause and Effect</strong>—Cause-effect determination is (strong ⭕) (medium 💧) (weak ⬠)</th>
<th><strong>Explanation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariation</td>
<td>⭕</td>
</tr>
<tr>
<td>Temporal precedence</td>
<td>⭕</td>
</tr>
<tr>
<td>No plausible alternative explanations</td>
<td>💧</td>
</tr>
</tbody>
</table>

---

**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 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.

---

<table>
<thead>
<tr>
<th>Pretest</th>
<th>Assignment</th>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_A$</td>
<td>C</td>
<td>1</td>
<td>X</td>
<td>$O_2$</td>
</tr>
<tr>
<td>$O_A$</td>
<td>C</td>
<td>2</td>
<td></td>
<td>$O_2$</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Assignment</th>
<th>Treatment</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ($n = 48$)</td>
<td>SMAT</td>
<td>C</td>
<td>Remedial math program</td>
<td>SMAT</td>
</tr>
<tr>
<td>2 ($n = 44$)</td>
<td>SMAT</td>
<td>C</td>
<td>-</td>
<td>SMAT</td>
</tr>
</tbody>
</table>

**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**.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type I Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>355.021</td>
<td>3</td>
<td>118.340</td>
<td>73.920</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>30183.773</td>
<td>1</td>
<td>30183.773</td>
<td>18856.264</td>
<td>.000</td>
</tr>
<tr>
<td>precut</td>
<td>270.863</td>
<td>1</td>
<td>270.863</td>
<td>169.212</td>
<td>.000</td>
</tr>
<tr>
<td>group</td>
<td>83.433</td>
<td>1</td>
<td>83.433</td>
<td>52.122</td>
<td>.000</td>
</tr>
<tr>
<td>interact</td>
<td>.725</td>
<td>1</td>
<td>.725</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>140.864</td>
<td>88</td>
<td>1.601</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30679.658</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>495.885</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The interaction is nonsignificant, indicating the two groups do not differ with respect to the slope of the regression line.

*a. R Squared = .716 (Adjusted R Squared = .706)*
Rerun Ancova Including Parameter Estimates

**Analyze → General Linear Model → Univariate**

Remove *interact* from the Covariates box. Click **OK**.

---

### Tests of Between-Subjects Effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type I Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>354.296*</td>
<td>2</td>
<td>177.148</td>
<td>111.351</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>30183.773</td>
<td>1</td>
<td>30183.773</td>
<td>16972.854</td>
<td>.000</td>
</tr>
<tr>
<td>precut</td>
<td>270.863</td>
<td>1</td>
<td>270.863</td>
<td>170.259</td>
<td>.000</td>
</tr>
<tr>
<td>group</td>
<td>83.433</td>
<td>1</td>
<td>83.433</td>
<td>52.444</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>141.589</td>
<td>89</td>
<td>1.591</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30679.858</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>495.885</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a. R Squared = .714 (Adjusted R Squared = .708)*

### Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Std. Error</th>
<th>t</th>
<th>Sig</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>23.407</td>
<td>.702</td>
<td>33.439</td>
<td>.000</td>
<td>22.092 - 24.833</td>
</tr>
<tr>
<td>precut</td>
<td>1.124</td>
<td>.083</td>
<td>13.611</td>
<td>.000</td>
<td>.960 - 1.288</td>
</tr>
<tr>
<td>group</td>
<td>-3.105</td>
<td>.440</td>
<td>-7.242</td>
<td>.000</td>
<td>-4.059 - -2.311</td>
</tr>
</tbody>
</table>

The differences in the groups are statistically significant.

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
Separate the Regression Lines

Data → Split File

Select Analyze all cases, do not split groups. Click OK.

Create Scatterplot

Graphs → Legacy Dialogs → Scatter/Dot
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,
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).

### Regression-Discontinuity Design: Validity, Control, and Causal Inferences

<table>
<thead>
<tr>
<th>Internal Validity — Threat is (high ‡) (medium ↔) (low ↓)</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>History</td>
<td>↑ Inclusion of pretest</td>
</tr>
<tr>
<td>Maturation</td>
<td>↑ Inclusion of pretest</td>
</tr>
<tr>
<td>Testing</td>
<td>↑ Inclusion of pretest</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>↓ The instrument did not change between pre and posttest</td>
</tr>
<tr>
<td>Statistical regression</td>
<td>↑ Treatment and control are asymmetric</td>
</tr>
<tr>
<td>Attrition</td>
<td>↓ No participants dropped out following the pretest</td>
</tr>
<tr>
<td>Selection bias</td>
<td>↓ Pre-post relationship at cutoff for each group is factored in</td>
</tr>
<tr>
<td>Combination of selection and other treatments</td>
<td>↓ Pre-post relationship at cutoff for each group is factored in</td>
</tr>
<tr>
<td>Diffusion</td>
<td>↓ Conditions are independent unless cutoff is “fuzzy”</td>
</tr>
<tr>
<td>Special treatment</td>
<td>↓ No-contact control group</td>
</tr>
<tr>
<td>Sequencing effects</td>
<td>n/a Between-subjects approach</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistical Conclusion Validity — Threat is (high ‡) (medium ↔) (low ↓)</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low statistical power</td>
<td>↑ Requires roughly 2.75 times the participants than an equivalent 2-group pre-post design w/ random assignment</td>
</tr>
<tr>
<td>Assumption violation of statistical tests</td>
<td>↓ Cutoff criterion stable; the comparison group pretest variance and continuous pretest distribution confirmed</td>
</tr>
<tr>
<td>Error rate problem</td>
<td>↔ Combination of analysis</td>
</tr>
<tr>
<td>Restriction of range</td>
<td>↓ Treatment and control conditions</td>
</tr>
<tr>
<td>Extraneous variance in the experimental setting</td>
<td>↔ Strong levels of design control implemented</td>
</tr>
<tr>
<td>Inaccurate effect size estimation</td>
<td>↓ ANCOVA plus linear regression is the appropriate analysis</td>
</tr>
<tr>
<td>Variability in procedures</td>
<td>↓ Strong levels of design control implemented</td>
</tr>
</tbody>
</table>
**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).

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
<th>Assignment</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>O₁</td>
<td>Mₐ</td>
<td>1</td>
</tr>
</tbody>
</table>

| Time |

Note: Mₐ 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: μ₁ = μ₂ 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.

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment</th>
<th>Posttest</th>
<th>Assignment</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 54</td>
<td>Satisfaction training</td>
<td>Job satisfaction survey</td>
<td>Gender</td>
<td>Males (n = 24)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Females (n = 30)</td>
</tr>
</tbody>
</table>

Data titled ExPostFacto.sav

Note: To review the scoring parameters for the job satisfaction survey, see Spector (1997).
Independent Samples $t$ Test

Analyze $\rightarrow$ Compare Means $\rightarrow$ Independent Samples $t$ Test

The Levene statistic is significant, indicating homogeneity of variance is violated.

**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).

The $t$ statistic shows significance ($p < .05$).

Males on average scored 44 points higher than females.

Point Biserial Correlation Coefficient (Effect Size)

Graph $\rightarrow$ Legacy Dialogues $\rightarrow$ Scatter/Dot $\rightarrow$ 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' $g$.

\[
d = \frac{\left| \bar{x}_1 - \bar{x}_2 \right|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}
\]

\[
d = \frac{149.08 - 105.50}{\sqrt{\frac{(23)(33.015)^2 + (29)(23.847)^2}{24 + 30 - 2}}} = \frac{43.58}{28.27} = 1.54
\]

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 $\rightarrow$ Correlate $\rightarrow$ 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 ($r_{pb} = -.615; R^2 = -.379$).

<table>
<thead>
<tr>
<th><strong>Ex Post Facto Design: Validity, Control, and Causal Inferences</strong></th>
<th><strong>Level</strong></th>
<th><strong>Explanation</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal Validity</strong>—Threat is (high ↑) (medium ↔) (low ↓)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>↑</td>
<td>Inclusion of pretest</td>
</tr>
<tr>
<td>Maturation</td>
<td>↑</td>
<td>Inclusion of pretest</td>
</tr>
<tr>
<td>Testing</td>
<td>↑</td>
<td>Inclusion of pretest</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>↓</td>
<td>The instrument did not change between pre- and posttest</td>
</tr>
<tr>
<td>Statistical regression</td>
<td>↑</td>
<td>Cannot determine based on data</td>
</tr>
<tr>
<td>Attrition</td>
<td>↑</td>
<td>No participants dropped out following the pretest</td>
</tr>
<tr>
<td>Selection bias</td>
<td>↔</td>
<td>Only reporting gender differences and groups were not self-selected</td>
</tr>
<tr>
<td>Combination of selection and other treatments</td>
<td>↑</td>
<td>Researcher did not control implementation of program</td>
</tr>
<tr>
<td>Diffusion</td>
<td>↑</td>
<td>Groups were in contact during the implementation phase</td>
</tr>
<tr>
<td>Special treatment</td>
<td>↑</td>
<td>Unknown but is likely due to lack of design controls</td>
</tr>
<tr>
<td>Sequencing effects</td>
<td>n/a</td>
<td>Between-subjects approach</td>
</tr>
</tbody>
</table>

(Continued)
Ex Post Facto Design: Validity, Control, and Causal Inferences

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Conclusion Validity</strong>—Threat is (high †) (medium ↔) (low ↓)</td>
<td></td>
</tr>
<tr>
<td>Low statistical power</td>
<td>†</td>
</tr>
<tr>
<td>Assumption violation of statistical tests</td>
<td>↔</td>
</tr>
<tr>
<td>Error rate problem</td>
<td>↔</td>
</tr>
<tr>
<td>Restriction of range</td>
<td>↓</td>
</tr>
<tr>
<td>Extraneous variance in the experimental setting</td>
<td>†</td>
</tr>
<tr>
<td>Inaccurate effect size estimation</td>
<td>↔</td>
</tr>
<tr>
<td>Variability in procedures</td>
<td>†</td>
</tr>
<tr>
<td>Subject heterogeneity</td>
<td>↔</td>
</tr>
<tr>
<td>Unreliability of the measures</td>
<td>↓</td>
</tr>
<tr>
<td>Multiple comparisons and error rates</td>
<td>↔</td>
</tr>
</tbody>
</table>

**Control**—Design and statistical control is (strong †††) (medium ††) (weak †)

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manipulation</td>
<td>†</td>
</tr>
<tr>
<td>Elimination</td>
<td>†</td>
</tr>
<tr>
<td>Inclusion</td>
<td>†</td>
</tr>
<tr>
<td>Group or condition assignment</td>
<td>†</td>
</tr>
<tr>
<td>Statistical procedures</td>
<td>††</td>
</tr>
</tbody>
</table>

**Cause and Effect**—Cause–effect determination is (strong †††) (medium ††) (weak †)

<table>
<thead>
<tr>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariation</td>
<td>†</td>
</tr>
<tr>
<td>Temporal precedence</td>
<td>†</td>
</tr>
<tr>
<td>No plausible alternative explanation</td>
<td>†</td>
</tr>
</tbody>
</table>

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.
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

<table>
<thead>
<tr>
<th>Validity, Control, and Causal Inferences</th>
<th>Level</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal Validity</strong>—Threat is (high ↑) (medium ↔) (low ↓)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maturation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instrumentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attrition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combination of selection and other treatments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diffusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequencing effects</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Statistical Conclusion Validity—Threat is (high ↑) (medium ↔) (low ↓)

<table>
<thead>
<tr>
<th>Threat</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Low statistical power</td>
<td></td>
</tr>
<tr>
<td>Assumption violation of statistical tests</td>
<td></td>
</tr>
<tr>
<td>Error rate problem</td>
<td></td>
</tr>
<tr>
<td>Restriction of range</td>
<td></td>
</tr>
<tr>
<td>Extraneous variance in the experimental setting</td>
<td></td>
</tr>
<tr>
<td>Inaccurate effect size estimation</td>
<td></td>
</tr>
<tr>
<td>Variability in procedures</td>
<td></td>
</tr>
<tr>
<td>Subject heterogeneity</td>
<td></td>
</tr>
<tr>
<td>Unreliability of the measures</td>
<td></td>
</tr>
<tr>
<td>Multiple comparisons and error rates</td>
<td></td>
</tr>
</tbody>
</table>

### Control—Design and statistical control is (strong ⬆️) (medium ⬇️) (weak ⬇️)

<table>
<thead>
<tr>
<th>Control</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manipulation</td>
<td></td>
</tr>
<tr>
<td>Elimination</td>
<td></td>
</tr>
<tr>
<td>Inclusion</td>
<td></td>
</tr>
<tr>
<td>Group or condition assignment</td>
<td></td>
</tr>
<tr>
<td>Statistical procedures</td>
<td></td>
</tr>
</tbody>
</table>

### Cause and Effect—Cause-effect determination is (strong ⬆️) (medium ⬇️) (weak ⬇️)

<table>
<thead>
<tr>
<th>Cause and Effect</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariation</td>
<td></td>
</tr>
<tr>
<td>Temporal precedence</td>
<td></td>
</tr>
<tr>
<td>No plausible alternative explanations</td>
<td></td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>Item No.</th>
<th>Checklist Item</th>
<th>Reported on Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title and abstract</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1a</td>
<td>Identification as a randomised trial in the title</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>Structured summary of trial design, methods, results, and conclusions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(for specific guidance see CONSORT for abstracts)</td>
<td></td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background and objectives</td>
<td>2a</td>
<td>Scientific background and explanation of rationale</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>Specific objectives or hypotheses</td>
<td></td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial design</td>
<td>3a</td>
<td>Description of trial design (such as parallel, factorial), including allocation ratio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3b</td>
<td>Important changes to methods after trial commencement (such as eligibility criteria), with reasons</td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>4a</td>
<td>Eligibility criteria for participants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4b</td>
<td>Settings and locations where the data were collected</td>
<td></td>
</tr>
<tr>
<td>Section/Topic</td>
<td>Item No.</td>
<td>Checklist Item</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Interventions</td>
<td>5</td>
<td>The interventions for each group with sufficient details to allow replication, including how and when they were actually administered</td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td>6a</td>
<td>Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6b</td>
<td>Any changes to trial outcomes after the trial commenced, with reasons</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>7a</td>
<td>How sample size was determined</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7b</td>
<td>When applicable, explanation of any interim analyses and stopping guidelines</td>
<td></td>
</tr>
<tr>
<td>Randomisation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence generation</td>
<td>8a</td>
<td>Method used to generate the random allocation sequence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8b</td>
<td>Type of randomisation; details of any restriction (such as blocking and block size)</td>
<td></td>
</tr>
<tr>
<td>Allocation</td>
<td>9</td>
<td>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</td>
<td></td>
</tr>
<tr>
<td>Concealment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mechanism</td>
<td>10</td>
<td>Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions</td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding</td>
<td>11a</td>
<td>If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11b</td>
<td>If relevant, description of the similarity of interventions</td>
<td></td>
</tr>
<tr>
<td>Statistical methods</td>
<td>12a</td>
<td>Statistical methods used to compare groups for primary and secondary outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12b</td>
<td>Methods for additional analyses, such as subgroup analyses and adjusted analyses</td>
<td></td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant flow</td>
<td>13a</td>
<td>For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome</td>
<td></td>
</tr>
<tr>
<td>(a diagram is</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>strongly</td>
<td>13b</td>
<td>For each group, losses and exclusions after randomisation, together with reasons</td>
<td></td>
</tr>
<tr>
<td>recommended)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>Item No.</th>
<th>Checklist Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recruitment</td>
<td>14a</td>
<td>Dates defining the periods of recruitment and follow-up</td>
</tr>
<tr>
<td></td>
<td>14b</td>
<td>Why the trial ended or was stopped</td>
</tr>
<tr>
<td>Baseline data</td>
<td>15</td>
<td>A table showing baseline demographic and clinical characteristics for each group</td>
</tr>
<tr>
<td>Numbers analysed</td>
<td>16</td>
<td>For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups</td>
</tr>
<tr>
<td>Outcomes and estimation</td>
<td>17a</td>
<td>For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)</td>
</tr>
<tr>
<td></td>
<td>17b</td>
<td>For binary outcomes, presentation of both absolute and relative effect sizes is recommended</td>
</tr>
<tr>
<td>Ancillary analyses</td>
<td>18</td>
<td>Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory</td>
</tr>
<tr>
<td>Harms</td>
<td>19</td>
<td>All important harms or unintended effects in each group (for specific guidance, see CONSORT for harms)</td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limitations</td>
<td>20</td>
<td>Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses</td>
</tr>
<tr>
<td>Generalizability</td>
<td>21</td>
<td>Generalizability (external validity, applicability) of the trial findings</td>
</tr>
<tr>
<td>Interpretation</td>
<td>22</td>
<td>Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence</td>
</tr>
<tr>
<td>Other information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Registration</td>
<td>23</td>
<td>Registration number and name of trial registry</td>
</tr>
<tr>
<td>Protocol</td>
<td>24</td>
<td>Where the full trial protocol can be accessed, if available</td>
</tr>
<tr>
<td>Funding</td>
<td>25</td>
<td>Sources of funding and other support (such as supply of drugs), role of funders</td>
</tr>
</tbody>
</table>

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, nonpharmaceutical 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.
Appendix H

CONSORT 2010 Flow Diagram

Assessed for eligibility ($n =$)  
Excluded ($n =$)  
- Not meeting inclusion criteria ($n =$)  
- Declined to participate ($n =$)  
- Other reasons ($n =$)  

Randomized ($n =$)  

Allocated to intervention ($n =$)  
- Received allocated intervention ($n =$)  
- Did not receive allocated intervention (give reasons) ($n =$)  

Allocated to intervention ($n =$)  
- Received allocated intervention ($n =$)  
- Did not receive allocated intervention (give reasons) ($n =$)  

Lost to follow-up (give reasons) ($n =$)  
Discontinued intervention (give reasons) ($n =$)  

Lost to follow-up (give reasons) ($n =$)  
Discontinued intervention (give reasons) ($n =$)  

Analysed ($n =$)  
- Excluded from analysis (give reasons) ($n =$)  

Analysed ($n =$)  
- Excluded from analysis (give reasons) ($n =$)  


Note: MS-Word versions of the checklist and flowchart can be found at http://www.equator-network.org/reporting-guidelines/consort/.
References


References


Kemp, A., Quintana, D. S., Quinn, C., Hopkinson, P., & Harris, A. (2014). Major depressive disorder with melancholia displays robust alterations in resting-state


Vagle, M. D. (2014). *Crafting phenomenological research*. Walnut Creek, CA: Left Coast Press.


Author Index

Abrami, P., 40
Abrams, L. S., 186
Aceee, 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
Bazely, P., 220
Behling, S., 113
Berman, H., 166
Bernard, R. S., 107
Bingham, G. E., 183
Birs, 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
Bulte, 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
Cappeller, 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
Cheung, 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
Conner, 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
Author Index

Coryn, C. L. S., 220
Creswell, J. W., 22, 143–144, 153, 157f, 172, 174n, 177, 178, 179, 180, 201, 210n, 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
Elliot, M. S., 187–188
Ellison, N. B., 137
Emalinda, M., 144
Erdfelder, E., 220
Erdogan, Y., 130
Ergul, C., 37
Escanilla, A., 47
Eveldt-Dawson, K., 39

Farahmand, F. K., 113
Paul, F., 220
Fetterman, D. M., 154
Finchman, J. E., 11t
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., 55

Ghazarian, S. R., 131
Gibney, C., 193
Gier, V. S., 83
Giuffre, M., 121
Glaser, B. G., 145, 147, 148f
Glass, V. G., 61
Gломбiewski, 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
Harro, J. M., 47
Harris, A., 237
Harris, G. T., 243
Haskell, W. L., 48
Hasson, F., 213
Hausensteine, 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
Hinkelmann, K., 63
Hinnant, J. B., 131
Hirakata, P., 164
Hlavinkova, L., 238
Hoagwood, K., 180
Hofer, M., 236
Holbert, D., 122
Holst, 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
Jacoby, J., 3
Jain, D., 203
Jones, E. W., 184
Jones, M. A., 136
Jones, P., 51
Jones, S. R., 150
Josseison, R., 144, 161
Kabaca, T., 130
Kaiser, M. K., 254
Kaldijian, 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, A., 142
Kraemer, H. C., 220, 241
Kramarski, B., 42
Kreiner, D. S., 83
Kristufkova, Z., 238
Lachin, J. M., 17n
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
Lemesheow, 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., 17n
Maxin, J., 194
May, H., 71
Mazzi, M., 70
McAdams, D., 161
McCamaridge, 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., 179n, 180
Mulcachy, G. A., 166
Mullenix, J. W., 77
Mundia, L., 135
Nagel, T., 205
Nastasi, B. K., 179n
Natz-Gonzalez, A., 83
Neumann, M., 154
Niehaus, L., 179n, 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
Pack, J., 240
Palinkas, L. A., 20, 180
Park, S., 41
Park, Y. S., 239
Patton, M. Q., 20
Percy, C., 242-243
Petras, H. R., 203
Pillai, N. S., 77
Plano-Clark, V. L., 177, 178, 179, 180, 201, 219
Polkinghorne, D. F., 160
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
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
Schaefer, 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
Strassler, A., 145, 147, 148f
Stringer, E. T., 212, 214
Stufflebeam, D. L., 220
Stufkken, 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., 17n
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, 193n, 210n, 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 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
  - embedded designs, 214–216
  - participatory action research, 213
  - reference resources, 214
- Alpha value (α), 241–242
- Alternating-treatment design, 105
- 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
- Assumptions violations, threats to statistical conclusion validity, 11
- Attention control group, 18
- Attention threat to construct validity, 10
- Attrition, 8
- Audio data analysis, 322
- Authoethnography, 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
- Case study designs, 143–144, 229–230
  - embedded designs, 190, 193, 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

357
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, 11t
Compensatory rivalry, 11t
Conceptual definition, 6
Concurrent nested mixed methods design, 191n
Concurrent triangulation design, 181, 184n
Confirmability, 324
CONSORT 2010, 334-337
Constant comparative method, 150
Construct confounding, 10t
Constructivist design, 146, 150-151, 218
Construct validity, 9, 240-241
Solomon N-group design, 93
survey research, 133, 134
threats to, 10-11t, 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, 56
Credibility, 324
Criterion-related validity, 241
Criterion sampling, 20
Criterion variables, 5-6, 127, 129
Critical design, 153, 156-157, 161-162, 165f, 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, 10t
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, 8t
Dispersion statistics, 242
Double pretest and posttest control group design, 223, 225
Double pretest design, 33f
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, 12t
Eidetic reduction, 169
Elimination and control, 14
Embedded approach, 144, 189-190
action research and, 214-216
case study design, 190, 193f, 194-195, 229
embedded-correlational design, 190, 192f, 193-194
embedded-experiment design, 190, 191-193
Emergent design in mixed methods, 178
Emerging design in grounded theory, 146, 148, 149
Subject Index • 359

Epocbe, 173
Error rate problem, 12t
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, 10t
Experimenter expectancies, threat to construct validity, 10t
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, 9t, 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 (\(\Delta\)), 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, 148f, 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, 7t
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, 7t
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, 7–8t, 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 x 3, 252, 284–292
3 x 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, 7t
Mean, 242
Measurement, 30
Measurement reliability, threats to statistical conclusion validity, 12t, 133
Measures of central tendency, 242–243
Mediation, 128
threat to external validity, 9t
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), 231n
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, 10t
Mono-operation bias, 10t
Multilevel mixed methods design, 181, 187–188
Multinomial logistic regression, 128
Multiple baseline design, 105, 111–114
Multiple comparisons, threats to statistical conclusion validity, 12t
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, 165f, 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
Subject Index

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
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

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, 198f, 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
dependent variables, 5, 127
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 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
data analysis, 252–253
data analysis, control group design, 255–261
data analysis, 2-factor design, 261–268
double design, 223
equivalent designs, 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
repeated-measures 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, 325–325
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, 12r
Reactivity to assessment, 10r, 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, 12r, 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–xviii
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, 11t
Response rate, 134–135
Sample-related threats to external validity, 9t
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, 7t, 120
Selective coding, 148
Semi transcription, 323
Sensitization, 11t
Sequencing effects
  counterbalancing, 17, 57, 58, 254
  threats to internal validity, 7, 8t
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, 104t
  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, 8t
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), 284–290
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, 11–12t
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, 11t
Statistical regression, threat to internal validity, 7t
Statistical software packages, 220
Stimulus characteristics, threat to external validity, 9t
Storytelling, 160, 331. See also Narrative perspective
Stratified random assignment, 15, 17f
Stratified sampling, 20
Subject heterogeneity effects, 12f
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–xviii
Testing threat to internal validity, 7t
Test-retest reliability, 239–240
Test sensitization, 11f
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, 11f
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, 10f
Treatment variations, threat to external validity, 9t
Trend analysis, 127
Trend-based data collection, 135
Triangulation design, 142
  concurrent design, 181, 184n

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, 60f, 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